TESIS DOCTORAL

MODELO DE HUMANOS PARA EL ESTUDIO DEL COMPORTAMIENTO HUMANO EN TAREAS DE SEGURIDAD Y VIGILANCIA

Encarnación Folgado Zúñiga

Licenciada en Ciencias Físicas,
especialidad de Fisica General

Departamento de Inteligencia Artificial
Escuela Técnica Superior de Ingeniería Informática
Universidad Nacional de Educación a Distancia

Madrid, diciembre 2012

Departamento de Inteligencia Artificial
Escuela Técnica Superior de Ingeniería Informática-UNED

MODELO DE HUMANOS PARA EL ESTUDIO DEL COMPORTAMIENTO HUMANO EN TAREAS DE SEGURIDAD Y VIGILANCIA

Autor Encarnación Folgado Zúñiga

Licenciada en Ciencias Físicas,
especialidad de Física General

Directores de la tesis: Dr. D. Mariano Rincón Zamorano

y Dr. D. José Manuel Cuadra Troncoso

TESIS DOCTORAL

MODELO DE HUMANOS PARA EL ESTUDIO DEL COMPORTAMIENTO HUMANO EN TAREAS SEGURIDAD Y VIGILANCIA

Autor Encarnación Folgado Zúñiga

Directores de la tesis: Dr. D. Mariano Rincón Zamorano

y Dr. D. José Manuel Cuadra Troncoso

Tribunal calificador

Presidente: Dr. D.

Secretaria: Dra. D$^{\underline{a}}$.

Vocal Primera: Dra. D$^{\underline{a}}$.

Vocal Segundo: Dr. D. e

Vocal Tercero: Dr. D.

Madrid, de de

Agradecimientos

En primer lugar, quisiera que mis palabras de agradecimiento fueran para todos los profesores del departamento de Inteligencia Artificial de la Escuela Técnica Superior de Ingeniería Informática de la UNED, que de una manera u otra me han ayudado en el transcurso de estos años. Especialmente quisiera hacer mención al profesor José Mira, el cual contribuyó al desarrollo de la Inteligencia Artificial en este país y que tuve el honor de poder conocer.

Quiero agradecer el apoyo y la ayuda prestada a mi tutor José Manuel de la Cuadra Troncoso del que he aprendido mucho. Pero de manera muy especial, quiero dar las gracias a mi tutor de tesis, profesor y amigo Mariano Rincón Zamorano por todo lo que ha hecho por mi durante todos estos años. No tengo palabras de agradecimiento para expresar la ayuda que me ha dado y que ha hecho posible que pueda presentar esta tesis, ya que sin él no hubiese podido hacerlo.

También quiero expresar mi enorme agradecimiento a mis padres, que durante años me han dado ánimos y apoyo para que siguiese estudiando.

A la memoria de mi hermano Antonio por el tiempo que pudimos estar juntos,

con todo mi cariño.

Resumen

Los sistemas de vigilancia se han utilizado durante muchos años como herramientas para monitorizar la seguridad en entornos vigilados. Estos sistemas se orientaron inicialmente a la mera centralización de la información en un centro base (vídeo, sensores de puertas y ventanas, sensores de presencia, etc.), donde personal humano se encargaba de interpretar la situación y actuar convenientemente. Carecían, por tanto, de capacidad para interpretar por sí mismos los eventos y los comportamientos relevantes para la tarea presentes en la escena. Tras los atentados del 11-S en Estados Unidos, el interés por el tema de la vigilancia aumentó significativamente a nivel mundial, aumentando la inversión para definir sistemas de monitorización inteligentes, flexibles y escalables, que aumentasen la capacidad de análisis de la escena realizada por el propio sistema, bien para la interpretación y toma de decisiones de forma automática, bien para facilitar la labor del vigilante.

Centrándonos en el reconocimiento de la actividad humana en tareas de vídeovigilancia, las actividades se consideran como eventos complejos, los cuales son definidos por composición espacio-temporal de eventos más simples, éstos a su vez por otros eventos aún más simples y así sucesivamente, formando una jerarquía de eventos, hasta enlazar con eventos primitivos, los cuales se determinan a partir de cambios en los atributos visuales de los individuos presentes en la escena.

En esta tesis se presenta el modelo BB6-HM, un modelo para la descripción de humanos basado en bloques que permite monitorizar, en tiempo real y con una carga computacional mínima, una gran cantidad de eventos primitivos relacionados con humanos. El modelo BB6-HM está inspirado en las reglas de proporcionalidad utilizadas en Bellas Artes a la hora de estructurar el primer esbozo de una figura humana. BB6-HM divide verticalmente el blob correspondiente a un humano en seis regiones de la misma altura. Cada una de estas regiones queda delimitada por el rectángulo que la contiene, denominado "bloque". Esta división en bloques de la silueta del humano permite definir multitud de parámetros que describen el comportamiento dinámico del humano y que pueden ser utilizados posteriormente para el reconocimiento de situaciones y eventos primitivos. El reconocimiento de tales eventos y situaciones se realiza, bien a través del análisis de los parámetros del modelo, bien de manera automática mediante la construcción de modelos, tanto estáticos como dinámicos, ajustados mediante aprendizaje automático supervisado. Estos eventos primitivos pueden utilizarse, a modo de librería, para la descripción de actividades más complejas en multitud de tareas de vigilancia (vigilancia de equipajes en aeropuertos, vigilancia de comportamientos de clientes en bancos, vigilancia de pacientes en hospitales, vigilancia de personas dependientes, etc.).

Abstract

Surveillance systems have been used for many years as tools to monitor and maintain security in supervised environments. Those systems were aimed initially at the mere centralization of information at a control room (video, window and door sensors, presence sensors, etc.), where human personnel was responsible for interpreting the situation and act accordingly. Therefore, those systems lacked of the capability to interpret by themselves events and situations on the scene relevant to the surveillance task. After the attacks of 11-S in the U.S., interest in the topic of surveillance increased significantly worldwide, increasing investment in intelligent monitoring systems to provide scene analysis enhanced capabilities, either to the automatic interpretation and decision making, either to facilitate the work of the security guard focusing her attention.

Focusing on human activity recognition in video-surveillance tasks, activities are considered as complex events, which are defined by spatio-temporal composition of simpler events, and these again for other events even simpler and so on forming a hierarchy of events, to connect with primitive events, which are determined from changes of state of visual attributes related to individuals in the scene and that can be obtained from video images.

This thesis presents BB6-HM model, a block-based model for the description of dynamic human behavior that can monitor, in real time and with a low computational load, many primitive events related humans. The model BB6-HM is inspired by the proportionality rules used in Fine Arts when an artist outlines a human figure. BB6-HM divides vertically the human blob in six regions of the same height. Each of these regions is bounded by the rectangle or bounding box, named "block". This division of the human silhouette into blocks will enable to define a multitude of parameters describing the dynamic behavior of humans that will be used later for recognizing primitive situations and events. The recognition of such events and situations is either done through the analysis of the model parameters, or automatically by building models, both static and dynamic, by supervised machine learning. These primitive events can be later used, as a library, to describe more complex activities in multiple surveillance environments (surveillance of baggage at airports, customer behavioral surveillance, hospital patient monitoring, monitoring of dependent people, etc.).

Índice general

1. Introducción general y resumen **1**
1.1. Motivación 1
1.2. Objetivos 2
1.3. Organización de la tesis 3

2. Estado del arte **5**
2.1. Modelado de humanos orientado a vigilancia 6
2.1.1. Modelado estático de humanos 6
2.1.2. Modelado del movimiento humano 9
2.1.2.1. Humanos lejanos: modelado con bajo nivel de detalle 12
2.1.2.2. Humanos cercanos: modelado detallado del humano completo 13
2.1.2.3. Humanos junto a la cámara: modelado de partes del cuerpo. 18
2.1.3. Modelos de humanos en sistemas de vigilancia 22
2.2. Reconocimiento de actividades 31
2.3. Reconocimiento de personas por la forma de andar (Gait) 35
2.4. Métodos de aprendizaje utilizados en la tarea de reconocimiento de actividades y eventos 37
2.4.1. Clasificadores basados en modelos estáticos 37
2.4.2. Clasificadores basados en modelos dinámicos 40

3. Descripción del modelo de humanos BB6-HM **41**
3.1. Contexto de la tesis propuesta 42
3.2. Descripción del modelo BB6-HM 44
3.2.1. Parámetros primarios y puntos significativos 48
3.2.2. Parámetros secundarios 48
3.3. Información proporcionada por el modelo 51
3.4. Localización de las partes del cuerpo 53

3.5. Información sobre el movimiento del humano 57
3.5.1. Análisis de la velocidad y la dirección del individuo 57
3.5.2. Análisis de la periodicidad 60
3.5.3. Análisis del ángulo del individuo respecto a la cámara 62
3.6. Reconocimiento de eventos y situaciones primitivas 63
3.6.1. Agacharse 64
3.6.2. Levantarse 65
3.6.3. Saltar (verticalmente) 66
3.6.4. Brazos levantados 66
3.6.5. Brazos extendidos 69
3.6.6. Tumbarse 70
3.6.7. Giros 70
3.6.8. Recoger objeto 71
3.6.9. Dejar objeto 74
3.6.10. Llevar objeto en la mano 74
3.7. Detección de oclusiones 78
3.8. Reconocimiento de personas por la forma de andar (Gait) 85

4. Experimentos **89**
4.1. Herramientas utilizadas 91
4.1.1. Imágenes utilizadas para la evaluación del modelo BB6-HM . 91
4.1.1.1. Base de datos de la UNED 91
4.1.1.2. Base de datos de CASIA 91
4.1.1.3. Preprocesamiento de imágenes 93
4.1.2. Librerías utilizadas 94
4.1.2.1. WEKA 94
4.1.2.2. Toolbox para los modelos ocultos de Markov (HMM's) 95
4.2. Experimentos realizados por métodos heurísticos 95
4.2.1. Experimentos sobre localización de partes del cuerpo 96
4.2.2. Experimentos sobre obtención de información del movimiento 96
4.2.3. Experimentos sobre reconocimiento de situaciones y eventos primitivos 99
4.2.4. Experimentos sobre detección de oclusiones 99
4.3. Experimentos realizados mediante aprendizaje supervisado 101
4.3.1. Evaluación de la orientación con respecto a la cámara 103
4.3.2. Reconocimiento del evento llevar objeto 105
4.3.3. Reconocimiento de la apariencia según la indumentaria: Normal (NM), Lleva Bolso (BG), Lleva Abrigo (CL) 105

4.4. Reconocimiento del "Gait" de personas 106
4.4.1. Reconocimiento del "Gait" de personas mediante modelos estáticos 107
4.4.2. Reconocimiento del "Gait" de personas mediante modelos dinámicos (modelos ocultos de Markov) 110
4.5. Resumen de los experimentos realizados 119

5. Aportaciones y trabajo futuro 123

6. Conclusiones 125

7. Anexo I 149
7.1. Herramienta WEKA 149
7.1.1. Algoritmos utilizados 150
7.1.1.1. Algoritmo J48 151
7.1.1.2. Algoritmo de Bagging 152
7.1.1.3. Algoritmo de boosting 153
7.1.1.4. Algoritmo de Stacking 153
7.1.1.5. Algoritmo MLP 154
7.1.1.6. Algoritmo SVM 155

8. Anexo II 157
8.1. Paquete Murphy Toolbox (HMM) 157
8.1.1. Funciones de entrenamiento en tiempo discreto 157
8.1.2. Funciones de entrenamiento en tiempo continuo 158
8.1.3. Funciones de clasificación en tiempo discreto 160
8.1.4. Funciones de clasificación en tiempo continuo 160
8.1.5. Funciones de generación de observaciones a tiempo discreto . . 161
8.1.6. Funciones de generación de observaciones a tiempo continuo . 162
8.1.7. Funciones de generación de la secuencia más probable 163

Nomenclatura

$BB6 - HM$ Block-based-six Human Model (Modelo Humano De Seis Bloques), página 42

CM Centro de Masas de la silueta, página 48

CW_{Wi}^{T} Change In Width (Vector de Cambio en la Anchura), página 50

DS_i Directional Symmetry Vector. (Vector De Simetría Direccional), página 50

HC Relación de altura de la entrepierna, página 50

H_T Altura del conjunto de bloques, página 48

P_C Punto superior del blob, el cual limita la altura del conjunto de bloques, página 48

P_Λ El punto de unión de las piernas en la silueta, página 48

P_{inf} Punto inferior del blob, el cual limita la altura del conjunto de bloques, página 48

S_f Swinging Feet Coefficient. (Coeficiente De Balanceo De Pies), página 51

S_h Swinging-hands Coefficient. (Coficiente De Balanceo De Manos), página 51

S_i Symmetry (Vector de Simetría), página 50

W_{B_i} Ancho de cada bloque i, página 48

W_{L_i} Porción de bloque a la izquierda desde el eje de simetría, página 48

W_{R_i} Porción de bloque a la derecha desde el eje de simetría, página 48

W_T Anchura máxima del conjunto de bloques, página 48

α	El ángulo que forman las manos respecto a la horizontal, página 48
θ	Carriyng Objects. (Coeficiente de Acarreo de Objetos), página 48
$\triangle CM_M^T$	Change In Mass Centre (Vector de Cambio en el Centro de Masas), página 50

Índice de figuras

2.1. Movimiento humano. Secuencias del tipo de movimiento. 12
2.2. Secuencia en la que un humano se introduce en un bosque y se aleja . A medida que se aleja, el blob se aproxima cada vez más a un punto. 13
2.3. Ajuste a la silueta obtenida del humano mediante elipsoides 3D. . . . 14
2.4. Modelo esqueletal a partir de un cuerpo en movimiento. Relación de puntos. 15
2.5. Modelo en estrella de Fujivoshi con cinco puntos. 16
2.6. Modelo de movimiento "Ribbons".(a) Vista Frontal (b) Vista Lateral. 16
2.7. Coordenadas del cuerpo humano en 3D y modelo de cilindros generalizados asociado. 17
2.8. Secuencia de humano en movimiento para la obtención del modelo de cilindros generalizados. 17
2.9. Representación del modelo basado en esferas. 18
2.10. (a) Secuencia original, (b) Secuencia con modelo elipsoide. 19
2.11. Modelo cilíndrico de la cara. 20
2.12. Modelo de la mano con sus uniones. 20
2.13. Modelo esqueletal de la mano. 21
2.14. Modelo de brazo con conos. 21
2.15. Frame de una secuencia de un lugar video-vigilado 24
2.16. Selección del contorno del rostro y partes del mismo para reconocimiento en tareas de vigilancia 27
2.17. Aplicación de sistema de vigilancia a humanos en el metro. Extracción de la silueta de personas en el metro para su tratamiento. 28

3.1. Esquema general del sistema 43
3.2. Tareas dentro del contexto de la vídeo-vigilancia 44
3.3. Estructura bottom-up detallada 45
3.4. Hombre de Vitruvio 45
3.5. Modelo de bloques en vista frontal y lateral. 47

3.6. Detalle del modelo BB6-HM. Puntos significativos y parámetros primarios del modelo BB6-HM. 49
3.7. Diagrama del sistema para reconocer eventos primitivos. 52
3.8. Ejemplo de eventos generados a partir de la secuencia para dejar un objeto. 54
3.9. Secuencia usada para la localización de las partes del cuerpo. Los puntos PH_D, PH_1, PH_2, PF_1, PF_2 y P_Λ están marcados con *. . . . 56
3.10. Coordenadas para la determinación del ángulo θ 59
3.11. Ángulo entre cabeza y pies según el desplazamiento. 59
3.12. Periodicidad del movimiento. 61
3.13. Ciclo del paso de dos humanos diferentes. 61
3.14. Gráfica de la periodicidad del movimiento. 62
3.15. Ángulo de una persona respecto a la cámara. 63
3.16. Secuencia de frames realizando la acción de agacharse. 64
3.17. Frame segmentada con el humano realizando la acción de saltar. . . . 67
3.18. Secuencia con la acción de elevar los dos brazos. 68
3.19. Acción de extender los brazos. 70
3.20. Puntos de los pies según la orientación. 71
3.21. Acción de recoger objetos. 73
3.22. Situación de llevar objeto en diferentes casos. 74
3.23. Situación de dejar un objeto. 76
3.24. Situación de objeto dejado. 76
3.25. Coordenadas para la determinación del ángulo γ 77
3.26. Variación del coeficiente Carriyng Objects cuando se lleva un objeto. 78
3.27. Variación del Carriyng Objects cuando no lleva un objeto. 79
3.28. Coeficiente HC cuando no lleva objeto o lleva de lado. 79
3.29. Asimetría y simetría del humano cuando lleva o no objeto. 80
3.30. Humano fuera de la oclusión. 81
3.31. Humano en la oclusión. 81
3.32. Representación de los puntos para la oclusión. 83
3.33. Oclusión parcial con el cuerpo obstaculizado a la mitad. 83
3.34. Casos correctos e incorrectos de la oclusión parcial. 84

4.1. Evaluación general de los experimentos realizados 90
4.2. Ejemplos de siluetas de la base de datos de CASIA 92
4.3. Imágenes de la base de datos de CASIA. 94
4.4. Diagrama del proceso para las reglas heurísticas. 97
4.5. Porcentaje de éxito por eventos analizados. 100

4.6. Matriz de confusión obtenida con WEKA. 104
4.7. Evolución temporal del parámetro HC 106
4.8. Resultados de los experimentos mediante aprendizaje supervisado. . . 111
4.9. Resultados medios de reconocimiento del "Gait" en función de los parámetros Q e It 113
4.10. Resultados por persona 117
4.11. Resultados algoritmo HMM. 118

7.1. Selección del algoritmo de J48 en Weka. 151
7.2. Selección de los Metabuscadores. 153
7.3. Selección del perceptrón multicapa (MLP) en Weka. 154
7.4. Selección de la máquina de soporte vectorial (SVM) en Weka. 155

Índice de tablas

2.1. Resumen del estado del arte en modelado en humanos 10
2.2. Resumen de estado del arte en Modelado de Humanos. 23
2.3. Publicaciones científicas más relevantes en vigilancia por número de citas. 30
2.4. Número de publicaciones relacionadas con la tecnología biométrica entre (2000-2010) 31
2.5. Enfoques utilizados para reconocimiento de actividades. 35

3.1. Posición habitual de las principales partes del cuerpo. 56

4.1. Porcentaje de localizaciones de partes del cuerpo. 98
4.2. Resultados de la obtención de información del movimiento 98
4.3. Resultados del reconocimiento de situaciones primitivas 99
4.4. Resultados de la detección de oclusiones 101
4.5. Tabla con los comandos de ejecución por métodos. 102
4.6. Tabla con los resultados de la evaluación de las personas 12, 15, 85. . 109
4.7. Tabla con los resultados según el ángulo de la cámara para los HMM's. 119
4.8. Comparativa entre los modelos estáticos. 120
4.9. Media de los resultados (estáticos y dinámicos), por ángulos. 120
4.10. Tabla con los resultados de la evaluación. 121
4.11. Tiempos medios de ejecución mediante reglas heurísticas. 122
4.12. Tiempos medios de entrenamiento y de test. 122

Capítulo 1

Introducción general y resumen

1.1. Motivación

Los sistemas de vigilancia se han utilizado durante muchos años como herramientas para monitorizar y mantener la seguridad en entornos sensibles. Estos sistemas se orientaron inicialmente a la mera centralización de la información en un centro base (video, sensores de puertas y ventanas, sensores de presencia, etc.), donde personal humano se encargaba de interpretar la situación y actuar convenientemente. Carecían, por tanto, de capacidad para interpretar por sí mismos los eventos y los comportamientos relevantes para la tarea de los individuos y objetos presentes en la escena. Tras los atentados del 11-S en Estados Unidos, el interés por el tema de la vigilancia aumentó significativamente a nivel mundial, aumentando la inversión para definir sistemas de monitorización inteligentes, flexibles y escalables, que aumentasen la capacidad de análisis de la escena realizada por el propio sistema, bien para la interpretación y toma de decisiones de forma automática, bien para facilitar la labor del vigilante.

Además, en los últimos años, dado el abaratamiento de la tecnología de captación de imágenes y su procesado, están apareciendo nuevos campos de aplicación para la interpretación de actividades y comportamientos también muy interesantes, como son el mundo de los juegos de ordenador, la supervisión de personas dependientes (tema de gran sensibilidad, dada la tendencia hacia una sociedad envejecida) o el estudio del comportamiento humano. Por tanto, la supervisión de personas y el reconocimiento automático de sus actividades tiene un gran interés desde distintos puntos de vista, no sólo interés económico o para la seguridad, sino también interés social pues, por ejemplo, puede permitir extender los años de independencia de las personas mayores, aumentando de este modo su calidad de vida.

Dentro del campo de la vigilancia, la monitorización de las actividades reali-

zadas por seres humanos es uno de los puntos de mayor interés actualmente. De hecho, resulta una tarea extremadamente compleja dadas las características de los actores presentes en la escena (objetos deformables, alta variabilidad en cuanto a su apariencia, multitud de comportamientos, coreografías complejas, etc.).

Por otro lado, desde el punto de vista operativo, los sistemas de seguridad deben ofrecer respuestas en tiempo real. Este requisito es primordial, pues el objetivo es detectar las situaciones de alarma lo antes posible y poder reaccionar a tiempo ante ellas. Otro tipo de sistemas, en los que se realiza un análisis forense de la situación, no presentan esta restricción. Por tanto, es necesario encontrar un equilibrio entre calidad del análisis y tiempo de respuesta. Por poner un ejemplo, existen algoritmos que permiten un seguimiento muy robusto de un individuo pero que requieren demasiada información de entrada o demasiada capacidad de cómputo, lo que les hace no ser adecuados para su integración en sistemas de vigilancia.

Por todo lo anterior, resulta de gran interés la investigación en sistemas de vigilancia inteligentes, robustos y que trabajen en tiempo real, para la monitorización de personas. Las aplicaciones de este tipo de sistemas son muy variadas: desde la vigilancia de infraestructuras por razones de seguridad, hasta la monitorización de personas dependientes (personas mayores en casa, enfermos eh hospitales, niños en guarderías etc.). Entre las tareas comunes a realizar se encuentran: 1) detectar la presencia de un humano en una secuencia de video; 2) identificar a un individuo entre un conjunto de personas o verificar la identidad de una persona y 3) caracterizar el comportamiento de un cierto individuo.

Teniendo en cuenta que la imagen es el sensor que ofrece mayor información de una determinada situación, el propósito de esta tesis será la caracterización del humano en secuencias de video: detección, identificación y reconocimiento y caracterización de la actividad realizada. Este tema se considera especialmente relevante, pues va en la línea de proporcionar una base sólida para la construcción de sistemas de interpretación de actividades humanas, los cuales pueden tener un enorme impacto en la sociedad actual y puede abrir la puerta a nuevos proyectos de gran impacto económico y social dentro de una sociedad con fácil acceso a la tecnología y altamente envejecida hacia la que se tiende.

1.2. Objetivos

Teniendo en cuenta los problemas planteados anteriormente, surgen una serie de objetivos concretos. El objetivo global es plantear un nuevo modelo de humanos que se pueda utilizar en tareas de vigilancia. De forma genérica, el modelo debe

permitir abordar, de manera robusta y en tiempo real, las tareas de detección e identificación de un individuo y el reconocimiento de la actividad que está realizando. Las actividades que puede realizar un humano son muy complejas y, en muchos casos, son coreografías en las que están implicados varios individuos (abrazarse, luchar, saludar, etc.), por lo que nos limitaremos al reconocimiento de eventos simples, relacionados con características visibles en la escena (de bajo nivel de abstracción) y que implican a un sólo humano (camina, corre, se sienta, coge una mochila, ...). Los comportamientos se considerarán eventos más complejos que se podrán describir como composiciones de estos eventos más simples.

Se distinguen los siguientes objetivos parciales o subtareas para llevar a cabo la investigación:

1. Estudio y análisis del estado del arte de los sistemas de seguridad actuales, así como de las técnicas empleadas para el análisis de comportamientos a partir de la información procedente de distintos tipos de sensores, especialmente de video.

2. Propuesta de un modelo de humanos orientado a vigilancia teniendo en cuenta las consideraciones obtenidas del estudio bibliográfico.

3. Utilización del modelo propuesto para el reconocimiento de eventos simples de interés para la tarea de vigilancia y relacionados con características visibles en la escena.

1.3. Organización de la tesis

En primer lugar, como apartado introductorio, se realizará un estudio crítico de los sistemas de vigilancia de humanos actualmente utilizados con el objetivo, por un lado, de identificar sus puntos fuertes y débiles, y por otro, de hacer explícito el conocimiento utilizado, tanto de forma explícita como implícita, para la definición de estos sistemas. Para ello, en el capítulo primero se hará un estudio del estado del arte desde la perspectiva de la vigilancia, donde se analizan trabajos previos relacionados con el modelodo de humanos y de sus actividades, con el reconocimiento de personas por la forma de andar, o con los métodos de aprendizaje utilizados en la definición de dichos modelos. En concreto, se identificarán qué características de un humano y qué estados y eventos son interesantes para la tarea de vigilancia.

Basándonos en este estudio. en el capítulo segundo, se propondrá un modelo para la caracterización del humano que sea útil en tareas de vigilancia. Se describirán, por tanto, sus características espacio-temporales y visuales, tales como el tamaño, la

forma, el desplazamiento, etc. Una vez definido el modelo, éste se utilizará para definir un sistema capaz de reconocer en la escena las acciones y los eventos de interés para la tarea de vigilancia. En concreto, se utilizará el modelo para extraer información acerca del movimiento, la periodicidad, primitivas de actividades y eventos relacionados con la postura y movimientos del humano.

Finalmente, en el capítulo tercero, se realizará la evaluación del sistema. Para ello, primero se comentarán las herramientas utilizadas imágenes, librerías, etc. A continuación, se expondrán los resultados obtenidos para el reconocimiento de actividades y el reconocimiento de personas por la forma de andar. . Para algunas de las actividades y eventos sencillos y para ciertas situaciones bien acotadas será posible, analizando los parámetros del modelo de humanos, describir modelos basados en heurísticas que permitan detectarlas de forma automática. Sin embargo, para otras situaciones más complejas, como para el reconocimiento de personas por la forma de andar, será necesario desarrollar sistemas clasificadores basados en aprendizaje automático a partir de ejemplos. Utilizaremos dos tipos de sistemas clasificadores estático y dinámico. En el capítulo cuarto, se comentarán las conclusiones, así como, posibles líneas futuras de investigación.

Capítulo 2

Estado del arte

Dividiremos el análisis del estado del arte en varios apartados, tratando de analizar los distintos puntos de vista que se han tenido en cuenta en el desarrollo de esta tesis doctoral. El marco conceptual de partida lo proporcionaMartinez-Tomas et al. (2008), donde se distinguen claramente los distintos niveles de descripción de una escena hasta proporcionar una descripción orientada a la tarea de vigilancia. De forma resumida, podríamos decir que, como en cualquier sistema, en toda descripción es necesario describir previamente los elementos constituyentes. En el caso de la tarea de vigilancia, para proporcionar una descripción orientada a la tarea (una descripción de alto nivel en términos de "intrusos", "abandono de objetos", "tumultos", "merodeadores", "peleas", "accidentes", "maniobra peligrosa", etc), es necesario partir de los elementos constituyentes (humanos, vehículos, maletas, etc). De todos estos posibles elementos constituyentes, en esta tesis nos centraremos en la descripción del comportamiento de los humanos presentes en la escena. Nuestro objetivo, por tanto, será detectar comportamientos y situaciones primitivas realizadas por humanos, relacionadas directamente con patrones espacio-temporales describibles en términos de características visibles en la secuencia de imágenes o video. La hipótesis de trabajo es que, después, por composición de patrones espacio-temporales de estos comportamientos y situaciones básicos (primitivas), será posible describir comportamientos más complejos y más cercanos a la tarea de vigilancia.

La organización de este capítulo será el siguiente. En la sección 2.1, comenzaremos estudiando los modelos que se han utilizado para representar la evolución espacio-temporal de nuestros objetos de interés, los humanos. En este sentido, primeramente nos centraremos en el modelado de las características espaciales (sección 2.1.1) y después de las temporales (sección 2.1.2). En estas dos secciones ya se comienza a anticipar las actividades y situaciones primitivas que son de interés en vigilancia y que deberán ser detectadas por nuestro modelo. Pero será en las seccio-

nes 2.2 y 2.3 donde realizaremos un estudio más orientado en ellas. En la sección 2.2, se realizará un estudio de las actividades de interés para la tareas de vigilancia y, en la secc. 2.3, dedicaremos especial atención al reconocimiento de personas por la forma de andar. Por último, dado que la descripción de algunos comportamientos y situaciones resulta compleja de realizar de forma analítica, en esta tesis se utilizará aprendizaje automático supervisado en varias ocasiones, por lo que se realizará un breve estado del arte sobre los métodos de aprendizaje más utilizados en tareas de vigilancia (sección 2.4).

2.1. Modelado de humanos orientado a vigilancia

En primer lugar se estudia, desde un punto de vista estático, la capacidad de representación de características espaciales y de análisis de la apariencia de distintos modelos de humanos. Esto permitirá identificar su capacidad para localizar la posición de las distintas partes del cuerpo de un humano y su capacidad de razonamiento sobre éstas para identificar posturas, signos, etc. Posteriormente, introduciremos el dinamismo de la escena y analizaremos la capacidad de representación y análisis del movimiento de los distintos modelos, lo que nos permitirá identificar qué actividades y eventos podrán ser detectados.

2.1.1. Modelado estático de humanos

La manera más sencilla de representar a un humano es mediante un punto, el cual está asociado normalmente a su centro de masas. Este modelo sólo permite obtener información de la posición y será tratado en más detalle en el apartado 2.1.2.1, ya que es el modelo más simple que permite abordar tareas de seguimiento de una persona situada a larga distancia de la cámara. La segunda manera más sencilla de representar a un humano es mediante la caja rectangular que lo contiene y que es paralela a los ejes (denominada en inglés “bounding box”). Este modelo ha sido ampliamente utilizado en tareas de vigilancia por su simplicidad Delaitre et al. (2010); Girondel et al. (2006); Broggi et al. (2000); Sokolova & Fernández-Caballero (2013). Otras formas simples utilizadas como modelo de representación son las formas elípticas Nakazawa et al. (1998). Estos modelos, aunque proporcionan una información demasiado básica y de carácter global, resultan bastante útiles en tareas de vigilancia, pues son modelos con muy poca carga computacional y pocas restricciones de aplicación y permiten analizar información relacionada con el tamaño de los objetos y su evolución en el tiempo (distancia a la cámara, distinguir entre humanos y otro tipo de objetos, etc.). Además, aunque no permiten identificar fácilmente las diferentes partes del cuerpo, sí dan una idea aproximada de la posición

de las partes más externas del cuerpo (la cabeza arriba, los pies abajo y las manos en los lados). También se han utilizado como una representación intermedia para delimitar la posible región de interacción de una persona con agentes del mundo que le rodea Darrell et al. (1994).

También son modelos simples aquellos basados en regiones de color uniforme. Estos modelos se han aplicado, por ejemplo, a los deportes, ya que cada jugador se puede modelar dividiéndolo en un número de regiones y clasificando por colores predominantes. En Pascual et al. (2006), el modelo se divide en dos o más regiones y cada región representa una parte del uniforme del equipo (camiseta, pantalón, medias, ...).

Otro grupo de modelos caracterizan a los humanos por su silueta. Algunos utilizan snakes o contornos activos Niyogi & Adelson (1994); Kass et al. (1988), ya que, por su flexibilidad, resultan muy eficaces para detectar y seguir el contorno de un objeto no rígido o deformable. El inconveniente principal de estos modelos es su dependencia de la inicialización (se debe tener una idea aproximada de la forma del objeto) y de la correcta ponderación de los distintos factores o energías que participan en el ajuste del modelo. Se destacan los trabajos de Haritaoglu et al. (1998a,b), que presentan un sistema en tiempo real para estimar la postura del cuerpo en humanos y detectar partes del cuerpo en imágenes monocromáticas. Para ello, analiza la silueta para determinar la ubicación de las partes del cuerpo utilizando diferentes relaciones de proporción entre dichas partes con el fin de encontrar una representación en forma de regiones poligonales básicas.

Otra manera simple de representar a un humano es mediante puntos significativos. En concreto, Wren et al. (1997a) utiliza tres puntos para representar a un humano, uno para la cabeza y otros dos para las manos. Estos puntos se detectan a partir de blobs basados en características de similitud de color e información espacial en 2D. Trabajos posteriores, como los de Azarbayejani et al. (1997) y Andersen et al. (2001), continúan en la misma línea en el espacio 2D, pero en futuras versiones se expandirán hasta el mundo 3D Wren et al. (1997a)García-Rojas et al. (2008).

Una evolución de los modelos anteriores consiste en la representación del esqueleto, lo que permite ampliar su utilización a la estimación de la postura o pose. Entre estos modelos están los que utilizan transformaciones sobre el eje medio GBharatkumar et al. (1994) o transformaciones de la distancia Iwasawa et al. (1999), que permiten el análisis de posturas de humanos en movimiento. Estos últimos son más avanzados y permiten suprimir partes que no son de mucho interés en el análisis, centrándose en aquellas que pueden aportar la información que se precisa. Así, por ejemplo, si lo que se necesita es buscar el centro de masas del humano, los brazos y

las piernas carecen de interés. Otro modelo interesante es el propuesto por Fujiyoshi & Lipton (2004), que consiste en generar una estrella desde el centro de masas hacia las extremidades (pies y manos) y la cabeza. Estos modelos tienen la ventaja de permitir estimar de manera sencilla la pose, por el contrario, son demasiado simples, como ocurre en otros métodos ya comentados, para dar una información completa sobre las acciones.

Aumentando la información asociada al esqueleto con los ángulos entre segmentos obtenemos los modelos articulados. Deutscher et al. (2000) propone un modelo formado por 17 segmentos con 29 grados de libertad, donde cada segmento es un cono de sección elíptica. La principal contribución de este trabajo es el desarrollo de un filtro de partículas modificado para optimizar la búsqueda en el espacio de configuración de alta dimensión. Este filtro de partículas modificado lo denominó "annealed particle filtering". Kehl et al. (2005) proponen un modelo para estimar el cuerpo entero con 24 grados de libertad desde múltiples cámaras. Este modelo combina un muestreo estocástico del conjunto de puntos del modelo en cada iteración del algoritmo lo que le evita quedarse atrapado en mínimos locales. Permite reconocer actividades como bailar.

Existen también trabajos en 3D, como Chen & Lee (1992), donde se define un modelo de barras que contiene 17 segmentos y 14 articulaciones que representan las características de la cabeza, torso, cadera, brazos y piernas. En Gavrila & Davis (1996) se localiza el torso en primer lugar y se utiliza para restringir la búsqueda de las extremidades mediante la descomposición del espacio de estados. El problema de usar modelos articulados es la alta dimensionalidad de la configuración del espacio y por tanto el crecimiento exponencial en el coste computacional.

Con la mejora del equipamiento tecnológico, el uso de escáneres en 3D ha permitido obtener modelos más precisos del cuerpo. Por ejemplo, en Allen et al. (2002) se utilizan múltiples escaneados de una persona en diferentes posturas para generar el modelo, o en Thalmann & Seo (2004) se usan bases de datos escaneadas en 3D para hallar relaciones entre las medidas de una persona. Más recientemente, la cámara Kinect Smisek et al. (2011); Arici (2012); Soltani et al. (2012), que utiliza un modelo de rejilla de puntos en el infrarrojo para obtener información de rango, se ha utilizado para aplicaciones de media distancia (2-10 metros). El principal problema es precisamente su alcance limitado.

En resumen, los modelos citados previamente, o son demasiado simples para caracterizar actividades y eventos complejos, o son demasiado sofisticados y sus necesidades de cómputo les hace ineficaces en problemas en tiempo real. La Tabla 2.1 resume todo lo comentado anteriormente. Otras revisiones bibliográficas del mode-

lado de humanos desde la perspectiva de la vigilancia, tanto desde el punto vista del cuerpo entero, como sólo a través de los gestos las encontramos en Gavrila (1999); Kushwaha et al. (2012); Zhang et al. (2011); Trinh et al. (2011).

2.1.2. Modelado del movimiento humano

El ser humano posee una amplia libertad de movimientos, lo que le permite desplazarse y realizar distintas acciones simultáneamente, bien en solitario (como andar, correr, agacharse, saltar, etc.), bien interactuando con otras personas (como pelear, saludar, abrazarse, etc.). Para la tarea de videovigilancia son de interés dos niveles de detalle en el análisis del movimiento humano, uno menos exigente, que permita detectar distintas situaciones, y otro más preciso, para llegar a caracterizar las situaciones y distinguir, por ejemplo, entre situaciones normales y anormales o el individuo que realiza la acción. Se distinguen, a su vez, dos tipos de movimientos, *los relacionados con el movimiento global del humano* (desplazamientos, rotaciones, trayectorias, etc.), que interesará en tareas de vigilancia como control de presencia, vigilancia perimetral, detección de velocidades anormales, detección de trayectorias anormales, etc., y *los relacionados con el movimiento de partes del cuerpo*, asociadas a eventos simples que pueden estar relacionados con actividades y gestos realizados por el humano, como agacharse, saludar, etc.

En la sección anterior se estudiaron distintas formas de representar el cuerpo humano de una manera estática. Aquí, al añadir el dinamismo a las acciones, aparece el problema añadido del seguimiento o, en inglés, tracking. En el seguimiento se distinguen fundamentalmente dos problemas, el de la selección de los elementos de interés que se desea seguir y el de establecer la correspondencia entre dichos elementos a lo largo del tiempo (este problema también se conoce como emparejamiento o, en inglés, "matching"). Los métodos de tracking se pueden clasificar en dos grandes grupos, los que detectan el movimiento mediante diferencias entre frames consecutivas (sustracción del fondo) y los que buscan los objetos de interés mediante análisis de la correlación con un patrón previo (en inglés, "template matching").

Los elementos de interés entre los que se establece la correspondencia pueden ser genéricos y no utilizar conocimiento del dominio, por ejemplo, pueden ser pixeles, como en el caso del flujo óptico (algoritmo de tracking de Lucas-Kanade descrito en **?**), o regiones homogéneas, como en el caso del método meanshift (algoritmo de tracking Camshift Bradski (1998)). Pero también es posible utilizar conocimiento del tipo de objeto y de su movimiento en instantes anteriores para predecir el comportamiento futuro y facilitar la tarea de tracking. El uso de estimadores, como el filtro de Kalman, siguen esta línea. El filtro de Kalman, el cual ha sido ampliamente utilizado en problemas de seguimiento Branko et al. (2004); Pomares et al.

Referencia	modelo	Inconvenientes	Ventajas
Darrell	Caja	Excesivamente simple para tareas de vigilancia	Sencillo, poca carga computacional.
Nakazawa	Formas elípticas	Difícil de identificar diferentes partes del cuerpo debido a la forma.	Sencillo, poca carga computacional.
Wren	tres puntos (blobs)	La identificación de los blobs humanos resulta aún más complicada cuando los objetos son pequeños, tal y como sucede en las tareas de vigilancia.	Manejan oclusiones y obtienen datos más significativos en el análisis de la excena.
Azarbayejani	puntos (blobs)	La identificación de los blobs humanos resulta aún más complicada cuando los objetos son pequeños, tal y como sucede en las tareas de vigilancia.	Manejan oclusiones y obtienen datos más significativos en el análisis de la excena.
Andersen	puntos (blobs)	La identificación de los blobs humanos resulta aún más complicada cuando los objetos son pequeños, tal y como sucede en las tareas de vigilancia.	Manejan oclusiones y obtienen datos más significativos en el análisis de la excena.
Gavrila	Esqueleto	Demasiado simple en tareas de vigilancia, no da demasiada información sobre las acciones.	Fácil para la estimación de la pose.
Bharatkumar	transformación eje medio	Necesita una posterior transformación para convertirlo en in modelo esqueletal.	Fácil para la estimación de la pose.
Iwasawa	Transformaciones de las distancias	Necesita una posterior transformación para convertirlo en in modelo esqueletal	Fácil para la estimación de la pose
Fujiyoski y Lipton	Esqueleto con forma de estrella	Demasiado simple en tareas de vigilancia, no da demasiada información sobre las acciones.	Fácil para la estimación de la pose
Chen y Lee	Modelo de barras que contiene contiene 17 segmentos y 14 articulaciones	Demasiado simple en tareas de vigilancia, no da demasiada información sobre las acciones.	Manejan bien las oclusiones, estimación de la pose.
Niyogi y Adelson	Ejes del esqueleto	Demasiado simple en tareas de vigilancia, no da demasiada información sobre las acciones.	Fácil para la estimación de la pose.
Haritaoglou	Proporciones de regiones	Mucha carga computacional	Manejan oclusiones y obtienen datos más significativos en el análisis de la escena.
Deutscher [25]	Conos con sección elíptica	Mucha carga computacional	Manejan oclusiones y obtienen datos más significativos en el

Tabla 2.1: Resumen del estado del arte en modelado en humanos

(2002); H.Vold et al. (1997); Roesser (1975), nos da un estimador lineal insesgado que, además, es el de menor varianza de todos los estimadores posibles, con lo que el resultado puede ser considerado como el mejor dada la información disponible. Sin embargo, los resultados son muy pobres si las variables de estado no siguen una distribución gaussiana. Una solución ampliamente empleada es el filtro de partículas (también conocido como algoritmo de condensación en el caso de seguimiento de contornos Isard & Blake (1998)), que permite modelar sistemas dinámicos con distribuciones de densidad de probabilidad no gaussianas y soporta de manera natural el chequeo de múltiples hipótesis por su naturaleza probabilística. Además, estos métodos, computacionalmente costosos, se han beneficiado del procesamiento en GPUs Montemayor & Sanchez (2005). Como regla general, se podría decir que, cuanto más conocimiento del dominio se modela en la definición del tipo de objeto a seguir, más fácil es la tarea de tracking, pues el número de elementos de interés disminuye drásticamente y esto simplifica el problema de la correspondencia entre frames (vease, por ejemplo, Fujiyoshi & Lipton (1998)).

Los resultados del tracking se pueden almacenar como "patrones de movimiento", los cuales capturan el movimiento en la imagen asociado a la realización de una determinada acción, o como trayectorias de partes concretas de los objetos presentes en la escena. Estos resultados, para ser utilizados porteriormente en tareas de reconocimiento, pueden guardarse directamente en bases de datos de movimientos (como secuencia de imágenes o de puntos de las trayectorias) o utilizarse para modelar las regularidades de ciertas características visuales inherentes a cada tipo de movimiento.

A su vez, los movimientos pueden modelarse de forma estática, mediante firmas de movimiento que no tienen en cuenta la posible variación temporal a la hora de realizar la misma acción por distintos individuos o por el mismo individuo en distintas circustancias, o de forma dinámica, donde sí se tiene en cuenta la variable temporal, utilizando por ejemplo redes bayesianas dinámicas o modelos ocultos de markov. Los primeros pueden ser válidos para movimientos de duración corta, pero enseguida es necesario recurrir a métodos dinámicos. En los métodos estáticos, la clasificación se realiza frame a frame: el descriptor espacio-temporal centrado en la frame se compara con las firmas de las distintas acciones reconocibles previamente almacenadas y se selecciona aquella con máxima correspondencia o, en caso de que haya varias acciones con alta correspondencia, se toma el máximo durante una ventana de tiempo significativa. En el caso de los métodos dinámicos, se suele generar un modelo por cada una de las acciones y seleccionar aquel cuya respuesta mejor explica las observaciones.

Figura 2.1: Movimiento humano. Secuencias del tipo de movimiento.

En las siguientes secciones se realiza un análisis bibliográfico de métodos para el modelado del movimiento de humanos con especial énfasis en el modelado de acciones relacionadas con la tarea de vigilancia. Las características distinguibles en la imagen y, por tanto, las posibilidades de utilizar conocimiento del dominio varían significativamente en función de la distancia a la cámara, por lo que dividiremos el estudio en tres subapartados: larga, media y corta distancia. En cualquier caso, a pesar de los esfuerzos realizados en las últimas décadas, el tracking sigue siendo una tarea compleja y aun no resuelta de manera robusta y fiable. Los principales problemas a los que se enfrenta el tracking siguen siendo los cambios de escala, oclusiones, cruces, cambios de forma o de iluminación.

2.1.2.1. Humanos lejanos: modelado con bajo nivel de detalle

Cuando observamos objetos lejanos no es posible distinguir entre las distintas partes del cuerpo (brazos, piernas, torso, cabeza, etc). Lo único que percibimos es una forma con bajo nivel de detalle y que se acaba aproximando a un punto al aumentar la distancia. En estas circustancias, únicamente podemos realizar un seguimiento de los objetos presentes en la escena desde su aparición hasta su salida de la misma (fig.2.2) y describir al objeto por su trayectoria y por características básicas de su forma como son su altura y anchura. En el proceso de seguimiento, es posible determinar si cambia de velocidad o se mueve en una determinada dirección. Los eventos que se pueden reconocer son aquellos relacionados con: 1) el análisis de la trayectoria aproximada de los objetos, tanto individuales (entra, sale, aparece, desaparece, trayectoria normal, entra en zona prohibida, etc.) como de interacción dentro de grupos (entra en grupo, sale del grupo, abandona objeto, etc.); 2) las características básicas de altura y anchura del blob (humano tumbado o de pie, etc.); y 3) la combinación de ambas informaciones, que permite detectar eventos

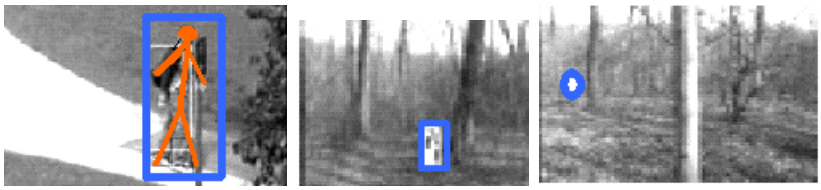


Figura 2.2: Secuencia en la que un humano se introduce en un bosque y se aleja . A medida que se aleja, el blob se aproxima cada vez más a un punto.

más abstractos, como si el objeto se acerca o se aleja, si se ha caído, si deja un objeto abandonado, etc.

Para objetos lejanos, el seguimiento de personas basado en regiones funciona bastante bien. En Fuentes & Velastin (2006), se describe un algoritmo basado en el emparejamiento bidireccional entre las cajas de dos frames consecutivas si éstas se solapan y que permite tratar con los problemas típicos del seguimiento de objetos (objetos que aparecen, desaparecen, forman grupos o se separan de un grupo). El algoritmo forma la trayectoria de los objetos a partir de la posición del centroide del blob a lo largo del tiempo y, a partir de esta trayectoria y de la forma de las cajas, detecta situaciones tales como personas que entran o salen de la escena, se alejan de la escena dejando algún equipaje, caídas, vandalismo, etc. Otro modo de realizar el emparejamiento entre blobs en frames consecutivas es utilizando redes neuronales, como en Do (2005), donde la red neuronal empleada es un perceptrón multicapa en secuencias de imágenes de baja resolución y utilizan la posición, la forma (altura y anchura de la caja) y el color como rasgos característicos de entrada a la red.

Tambien se han utilizado los contornos activos para el seguimiento Niyogi & Adelson (1994); Chenyang & Prince (1998); McInerney & Terzopoulos (1995); Zhang & Freedman (2005). Estos métodos permiten una asociación más precisa entre puntos concretos de los objetos en frames consecutivas pero, a larga distancia, esta no es una característica muy relevante. Además, computacionalmente son más costosos que el seguimiento basado en regiones y son altamente sensibles a la inicialización, haciéndose muy difícil la inicialización automática.

2.1.2.2. Humanos cercanos: modelado detallado del humano completo

En este apartado se trata con humanos que se mueven a una distancia media de la cámara. En estas circunstancias, sí podemos reconocer distintas partes del cuerpo humano y, por tanto, aplicar técnicas más complejas. Basándonos en los modelos de humanos descritos en la sección 2.1.1, el seguimiento se puede realizar a partir de representaciones que utilizan conocimiento a priori de la figura humana o sin conocimiento a priori. A continuación, se presentan algunos modelos y sus

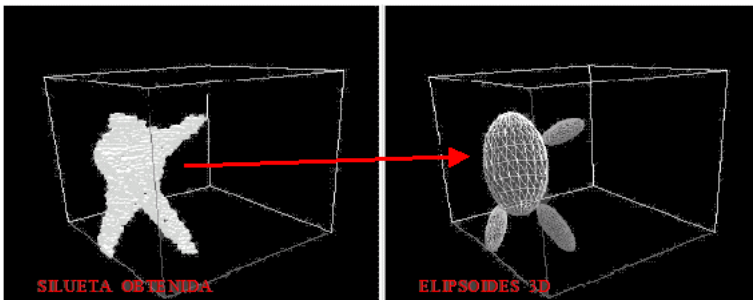


Figura 2.3: Ajuste a la silueta obtenida del humano mediante elipsoides 3D.

características más relevantes.

Este modelo utiliza elipsoides para aproximar las partes del cuerpo humano utilizando una única cámara o también múltiples cámaras Bregler & Malik (1998); Begler & Malik (1997). El movimiento lo describe en términos de un modelo de movimiento de la imagen centrado, mediante máscaras, en las partes del cuerpo humano. Permite una amplia libertad de movimientos a la hora de aproximar el cuerpo humano al modelo en secuencias complejas y son algoritmos que trabajan en tiempo real tras una inicialización manual.

Este modelo Cheung et al. (2000) utiliza seis elipsoides 3D para modelar las partes del cuerpo (cabeza, tronco, brazos y piernas) en un sistema multi-camara para el seguimiento de movimientos de humanos en tiempo real. El sistema consta de cinco cámaras, cada una conectada a un ordenador, el cual extrae localmente la silueta de la persona moviéndose en la imagen y la envía a un ordenador central donde se realiza una resonstrucción 3D basada en voxeles. Después, se realiza el ajuste a un modelo de seis elipsoides 3D (fig.2.3) mediante la segmentación por distancia a las elipses localizadas en la frame anterior y una posterior aproximación por análisis de momentos invariantes. La velocidad de trabajo es de más de quince frames por segundo.

El sistema “Pfinder” Wren et al. (1997a) es un sistema en tiempo real para el seguimiento de personas y la interpretación de sus conductas. Utiliza un modelo de representación multi-blob en el que los blobs se modelan por su posición y un modelo de mezcla de distribuciones gaussianas de color. A los blobs se les asigna parámetros estadísticos, tales como su posición media y su covarianza, y se incorpora conocimiento a priori para compensar pequeños cambios de iluminación no bruscos.

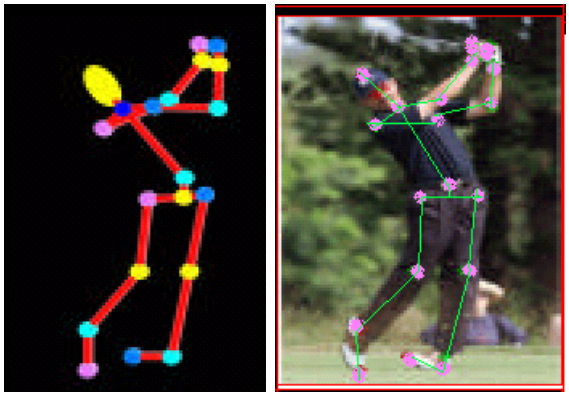


Figura 2.4: Modelo esqueletal a partir de un cuerpo en movimiento. Relación de puntos.

Realiza el análisis para un único humano en la escena Ogden & Dautenhahn (2001); Wren et al. (1997b) y presenta problemas para la asociación entre blobs y contornos, lo cual puede conducir, finalmente, a errores de clasificación y seguimiento que hagan que el sistema se vuelva inestable.

Modelos de humanos basados en HMM sobre modelos esqueletales Bobick & Wilson (1995)Wilson & Bobick (1999) aplican modelos ocultos de markov a modelos esqueletales para modelar propiedades espacio-temporales del movimiento humano (fig.2.4). Esta estructura se ha utilizado para analizar la dinámica del movimiento humano Bregler (1997), analizar los movimientos de humanos realizando distintas acciones, como correr o andar Krahnstover et al. (2001), o reconocer posturas por asociación a modelos HMM predefinidosChang & Huang (2000).

Modelos de humanos basados en HMM sobre modelos esqueletales Fujiyoshi & Lipton (1998) utiliza un modelo esqueletal en forma de estrella (fig.2.5). En esta propuesta, el tracking queda simplificado a la correspondencia de las distintas partes del cuerpo localizadas en cada frame. Dos tipos de movimientos se analizan utilizando este modelo de estrella, por un lado los posturales del cuerpo y por otro lado movimientos cíclicos de las diferentes partes del esqueleto Lipton et al. (1998). A partir de estos movimientos se determinan las actividades humanas, tales como andar, correr, etc.

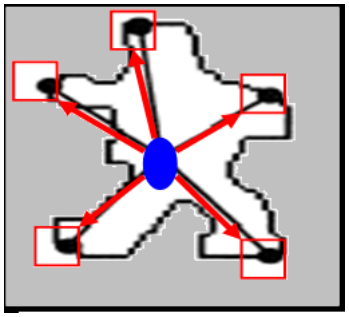


Figura 2.5: Modelo en estrella de Fujivoshi con cinco puntos.

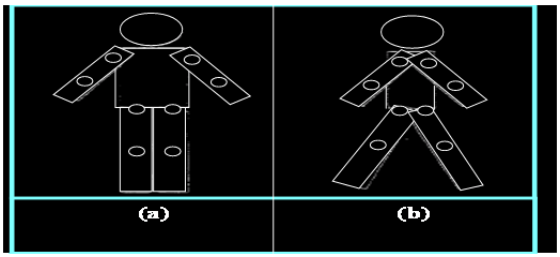


Figura 2.6: Modelo de movimiento "Ribbons".(a) Vista Frontal (b) Vista Lateral.

Modelo de humanos de Chang-Huang. Utiliza un modelo articulado Cheng et al. (1996, 1998) que consiste en identificar las extremidades del humano y calcular la variación del ángulo de las articulaciones. Utiliza un modelo de cuerpo en 2D con ocho uniones a considerar: hombros, codos, cadera y rodillas, (fig.2.6). Analizando el movimiento de cada una de las articulaciones, crea curvas paramétricas que describen los ángulos entre el eje principal del tronco con los brazos y las piernas. Estas curvas paramétricas se almacenan en una base de conocimiento y permiten comparar la situación actual con diferentes tipos de movimiento.

Modelo de humanos de Chang-Huang. Marr y Nishihara Marr & Nishihara (1978) utilizan el modelo de cilindros generalizados para representar al humano. El modelo propone un esquema completo para poder obtener una representación 3D de un humano. Para ello, añadieron información relativa a la profundidad y la orientación de las superficies y luego las agruparon en diferentes partes en 3D, dando lugar a formas cilíndricas (Figs. 2.7 y 2.8). En Marr & Nishihara (1975), cada cilindro se describe mediante los parámetros: radio largo, radio corto y altura. Se definen 25 parámetros de movimiento del humano, de los cuales 22 son parámetros angulares y 3 son de posición. Se asume un movimiento de 3 ángulos de rotación para la cabeza y dos ángulos para el torso.

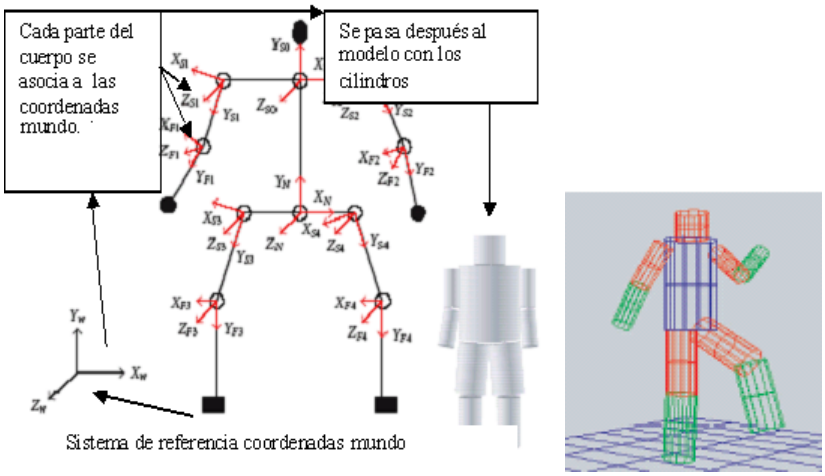


Figura 2.7: Coordenadas del cuerpo humano en 3D y modelo de cilindros generalizados asociado.

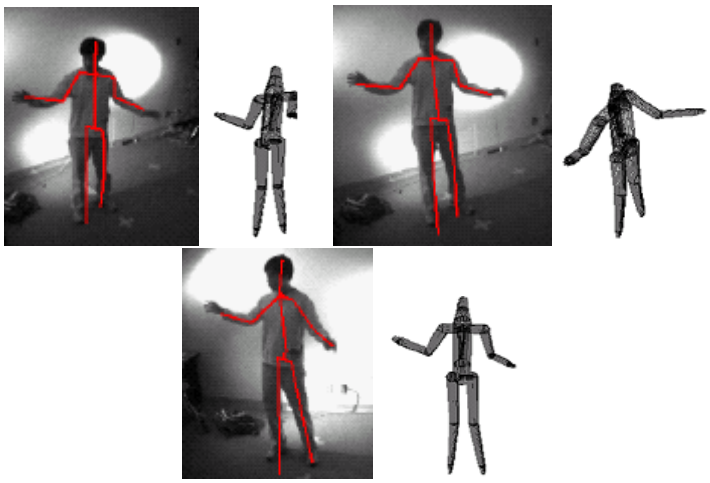


Figura 2.8: Secuencia de humano en movimiento para la obtención del modelo de cilindros generalizados.

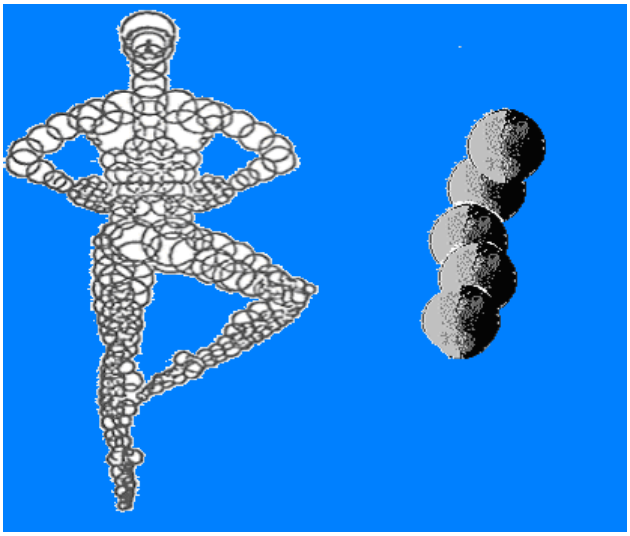


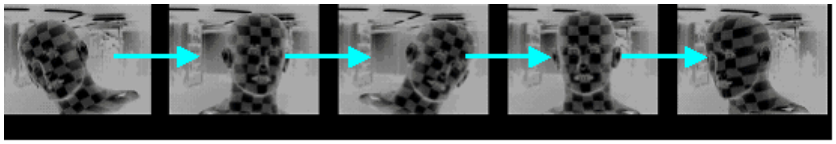
Figura 2.9: Representación del modelo basado en esferas.

Modelo de humanos de Chang-Huang. Este modelo Badler et al. (1979) utiliza esferas tridimensionales solapadas para representar el cuerpo humano y sus movimientos en tres dimensiones. El modelo consta de 310 esferas articuladas con 19 puntos de unión entre 20 diferentes segmentos (fig. 2.9). Los movimientos se definen en términos de cambios angulares entre articulaciones. El modelo está organizado en una estructura de árbol, con los segmentos como nodos y las uniones como arcos. El sistema de coordenadas es local a uno de los segmentos, el cual se convierte en el nodo raíz, lo que permite la visualización de los movimientos del cuerpo respecto a cualquier segmento de referencia. Por ejemplo, es posible analizar el movimiento de las piernas relativo a la columna vertebral. Una representación similar fue utilizada por Aswatha et al. (1996).

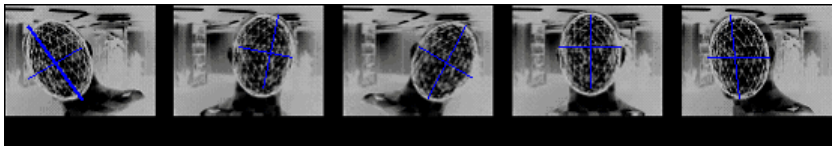
2.1.2.3. Humanos junto a la cámara: modelado de partes del cuerpo.

En este caso, la cámara está colocada muy cerca del humano, por lo que no se analiza al humano completo, ya que no interesa su desplazamiento sino únicamente los gestos y signos realizados, principalmente por la cara o por las manos. A continuación se describen algunos modelos.

Modelo de humanos de Chang-Huang. Basu et al. (1996); Essa et al. (1996) utilizan elipsoides 3D para modelar el movimiento de la cabeza con seis grados de libertad e interpreta el flujo óptico en términos de posibles movimientos rígidos del modelo. Este modelo ha sido aplicado a cabezas con formas y estilos de pelo diferentes, con resultados muy realistas al mejorar la calidad de la alineación de los puntos.



(a)



(b)

Figura 2.10: (a) Secuencia original, (b) Secuencia con modelo elipsoide.

Sin embargo, con el fin de generar una visión de la cara compatible con la rotación especificada de la cabeza, necesita la generación de una plantilla inicial, lo cual conduce a un tiempo de cómputo bastante elevado. El modelo de elipsoides es robusto a pequeñas variaciones en el ajuste inicial, permitiendo de este modo la inicialización automática del modelo. El seguimiento resulta bastante estable (fig. 2.10). Su uso se ha probado con diferentes números de "frames" obteniéndose resultados positivos de hasta con 30 "frames" por segundo con imágenes de muy alta calidad y de 5 "frames" por segundo para las de baja.

Modelo para seguimiento de cabeza de La Cascia, Isidoro y Sclaroff. Cascia et al. (1998) proponen un algoritmo para seguimiento de la cabeza en 3D basado en texturas mapeadas sobre un modelo de superficie 3D cilíndrico. Se puede usar para reconocimiento de labios u otro tipo de expresión, exigiendo, para esto, que la posición frontal de la cabeza. El modelo tiene dificultades cuando se efectúan rotaciones rápidas y largas de la cabeza. Permite el manejo de oclusiones. El control de movimientos resulta bastante mejor que utilizando un modelo de 2D, ya que el cilindro constituye una mejor aproximación a la cara que el plano (fig. 2.11).

Modelo para seguimiento de la mano de Cohen Este modelo se describe en Cohen (2004)Cohen & Lee (2002); Sung & Cohen (2004). Usa un modelo basado en cilindros y esferas para obtener una representación articulada en 3D para el modelado del movimiento de la mano. El modelo completo consta de 15 puntos de unión ("joints") y 20 grados de libertad, tres puntos de unión y una yema por cada dedo (fig.2.12). Los cilindros tienen un radio y una longitud ajustable y están conectados

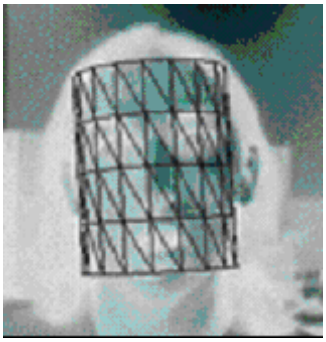


Figura 2.11: Modelo cilíndrico de la cara.

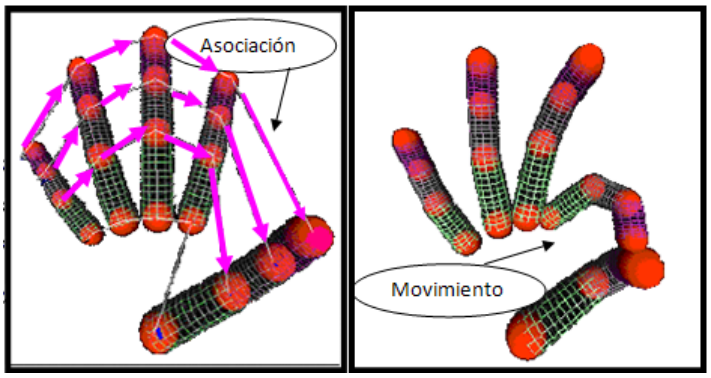


Figura 2.12: Modelo de la mano con sus uniones.

mediante puntos de unión. Todas las uniones tienen rotación y el movimiento de la mano ha sido modelado teniendo en cuenta el movimiento de los dedos y su relación con los dedos adyacentes. Este modelo se inicializa automáticamente para detectar la postura inicial y utiliza un filtro de partículas para ajustar y seguir el movimiento de las manos y los dedos.

Modelo de seguimiento de la mano de Wu, Lin y Huang Wu et al. (2001) utilizan un modelo esqueletal para representar la mano que se compone de un conjunto de clases de objetos interrelacionadas. Cada dedo es modelado como una cadena cinemática donde la palma es el sistema de referencia (fig.2.13) en la cual se introducen restricciones adicionales al movimiento de la mano como, por ejemplo, los ángulos máximos y mínimos para cada grado de libertad o que existen dependencias de movimiento entre las articulaciones. Esto es importante para reducir el espacio de búsqueda al aplicar cinemática inversa para inferir la postura de la mano a partir de la posición de los dedos. Para la descripción del comportamiento se realiza un estudio global del movimiento y, posteriormente, se focaliza en los movimientos

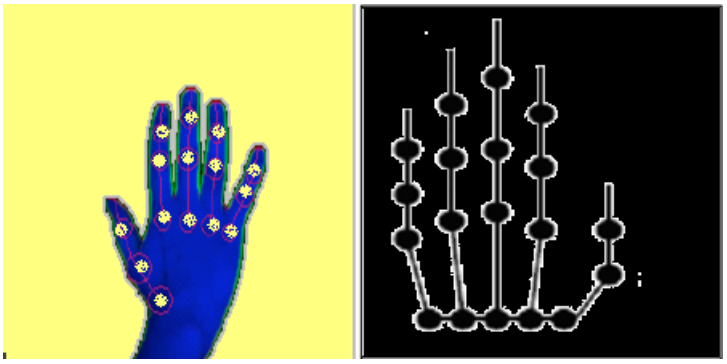


Figura 2.13: Modelo esqueletal de la mano.

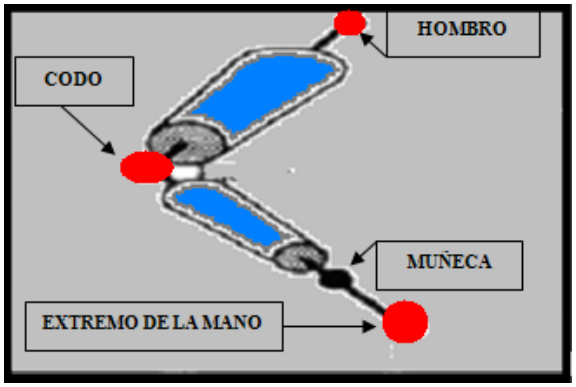


Figura 2.14: Modelo de brazo con conos.

locales de cada uno de los dedos (Lin et al. (2000)).

Modelo para seguimiento del brazo de Di Bernardo, Gonçalves, Peron y Ursella. Di Bernardo et al. (1996) proponen un modelo basado dos conos cilíndricos conectados mediante uniones esféricas para estudiar el movimiento de un brazo en 3D sin ningún tipo de restricciones (fig. 2.14). La idea es realizar un modelo del brazo en 3D y utilizar la estimación actual de la posición para predecir la proyección en la imagen. La diferencia entre la situación predicha y la actual es utilizada como medida de error para actualizar la posición estimada mediante un estimador recursivo, en concreto, un filtro de Kalman extendido. Así, en lugar de extraer características explícitamente de la imagen, realizan comparaciones directas entre la imagen actual y la esperada. Este modelo requiere de siete parámetros para describir la forma: longitud de las manos, antebrazos, diámetros de las articulaciones, etc. Para la inicialización se supone la posición del hombro conocida.

Como conclusiones de este estudio bibliográfico respecto del modelado de humanos, cabe destacar que, en el caso de que el modelado se realice con humanos

lejanos de la cámara, el estudio se ciñe fundamentalmente al tracking sobre el humano sin tener muchas más posibilidades de análisis en lo que se refiere a otro tipo de actividades. En caso del modelado de humanos cercanos a la cámara se han presentado diferentes tipos de modelos (elipses, cilindros, esferas, esqueletales, etc.) que proporcionan más información respecto a las acciones realizadas que en el caso anterior. Así, en el caso de utilizar los modelos esqueletales estos proporcionan buen rendimiento en cuanto al coste comutacional, sin embargo, las actividades humanas en vigilancia que permiten analizar son bastante simples. Por contra, si el tipo de modelo es más sofisticado (esferas, cilindros, etc), el estudio es más preciso y permiten, en general, obtener mejores resultados que los anteriores, sin embargo el coste computacional aumenta de manera exponencial. En cuanto al modelado de humanos junto a la imagen se centra en detectar situaciones más espefícicas, en general, el movimiento de alguna parte del cuerpo. Esto permite estudiar situaciones concretas (seguimiento cabeza, brazos, etc) pero no el conjunto de la actividad humana, además de que, al igual que en el caso anterior, requieren un alto coste computacional. En la Tabla 2.2, se muestra el resumen del estado del arte descrito en esta sección.

2.1.3. Modelos de humanos en sistemas de vigilancia

En los sistemas de vigilancia tradicionales es necesario que un operador supervise continuamente las cámaras que monitorizan un escenario desde la sala de control, lo cual supone una actividad bastante tediosa y que puede llevar a un pobre desempeño de la tarea por falta de atención o por sobrecarga de información. La fatiga visual, el elevado número de cámaras, el ritmo de los eventos que componen una actividad, el nivel de detalle de la observación, la disponibilidad horaria o el coste son sólo algunos de los factores que hacen pensar en el uso de un sistema de vigilancia automático. Los sistemas actuales tratan de incorporar un conocimiento específico sobre la forma humana y su apariencia para intentar proporcionar al vigilante únicamente información relevante o, en el caso extremo, una interpretación automática de la escena.

En la sección anterior se realizó una clasificación de modelos de humanos en función del nivel de detalle necesario para su aplicación. Ahora, nos centraremos en los sistemas de vigilancia que se han descrito en la bibliografía, los cuales estudian con mayor nivel de detalle las restricciones impuestas por los escenarios reales (tiempo real, cambios de iluminación, múltiples personas simultáneamente que forman grupos que cambian en el tiempo, etc.).

Así, F.Porikli & Tuzel (2003) presenta un sistema automático de seguimiento y

Descripción	Método	Tipos de Modelo	Ventajas	Inconvenientes	Algoritmo Matching	Actividades
MODELO BAJO HUMANOS LEJANOS DE LA IMAGEN	Métodos de tracking	Masa puntual	Son métodos que trabajan muy bien si no se produce ningún cambio de escala en el humano que se está siguiendo.	Se obtienen objetos muy lejanos, estos se acaban aproximando a una masa puntual. Presentan problemas en determinadas circunstancias, tales como, oclusiones, cruces, cambios de forma o de iluminación.	Filtro Kalman, Contornos Activos, Meanshift.	Tracking, entrar salir de la escena.
MODELO DETALLADO, HUMANOS CERCANOS A LA CAMARA	Modelos de aproximación de Bregler y Malik.	Elipsoides	Permite un amplio grado de libertad en secuencias complejas a la hora de aproximar el cuerpo humano al modelo y son algoritmos de alta velocidad.	Hay que inicializar el primer "frame" por parte del usuario. El proceso de reconocimiento incluye una actualización de "frame" a "frame."	Filtro Kalman.	Tracking.
	Modelos de aproximación de Cheung, Kanade, Bouguet y Holler.	Elipsoides 3D	Utiliza un sistema bastante iterativo que puede displayar y ser observado por un usuario en tiempo real.	Alto coste computacional.	Template matching.	Determinación postura, movimiento, análisis de rendimiento deportivo, juegos.
	Modelos de Bobick y Wilson.	Estados	Funcionan muy bien en reconocimiento dinámico de modelos humanos.	Alto coste computacional al incrementarse el número de imagenes.	Modelos de Markov.	Análisis de movimientos de humanos (correr o andar, gestos).
	Modelos de representación de Chang-Huang.	Modelo ribbons: cuerpo en 2D con ocho uniones del cuerpo	Este es un modelo más sofisticado, más avanzado que el modelo esqueleto o "stick model" y permite obtener mejores resultados que éste.	Demasiado simple en tareas de vigilancia, no da demasiada información sobre las acciones.	Template matching.	Determinación postura y análisis de movimiento.
	Modelo de movimiento de Fujiyoshi.	Modelo Esqueletal	Puede determinar actividades humanas con poco coste computacional.	Las actividades humanas determinadas son simples, correr y andar. No considera acciones para vigilancia.	Filtro Kalman.	Determinación postura, movimiento, análisis de movimiento, reconocimiento de acciones (pasear, correr).
	Modelo de movimiento de Wren. Pfinder.	Multiblob	Es un sistema en tiempo real para seguimiento de personas e interpretar sus conductas. El sistema puede compensar pequeños cambios de "iluminación.	Tiene alguna dificultad de integración entre blobs y contornos, lo cual puede conducir finalmente, a algunos errores de clasificación y seguimiento, que hagan que el sistema se convierta en inestable.	Filtro Kalman.	Aplicaciones en tiempo real (andar, sentarse, movimiento manos, gestos, etc).
	Modelo de movimiento de Marr y Nishihara.	Cilindros Generalizados	Completo esquema para poder obtener una representación 3D de un humano.	Alto coste computacional.	Correspondencia de la forma.	Determinación postura, movimiento, análisis de movimiento.
	Modelo de movimiento de O'Rourke y Badler.	Esferas tridimensionales	Las iteraciones entre el cuerpo y el entorno pueden detectarse y describirse. Preciso en el análisis del movimiento.	Alto coste computacional.	Template matching.	Analizar el movimiento de las piernas relativo a la columna vertebral.
MODELADO HUMANOS JUNTO DE LA IMAGEN	Modelo de movimiento de cabeza de Basu, Essa y Pentland.	Elipsoides 3D	Este método extrae exactamente los parámetros de movimiento en 3D. Su uso se ha probado con diferente número de "frames" obteniéndose resultados positivos de hasta con 30 "frames" por segundo.	Alto coste comutacional. A más de 30 frames por segel resultado se vuelve negativo.	Correspondencia de la forma.	Determinación postura cabeza.
	Modelo de movimiento de manos en 3D de Isaac Cohen.	Cilindros y Esferas	Buen resultado en el seguimiento de las manos y se inicializa automáticamente.	Alto coste computacional. A más cilindros y esferas definidas el modelo se vuelve cada vez más inoperativo.	Filtro de condensación.	Determinación postura de la mano.
	Modelo de movimiento de manos de Lin, Wu y Huang.	Modelo esqueletal de mano	Puede determinar variación de movimiento en las manos con poco coste computacional.	Limitado en cuanto a la variedad de acciones a tratar. Restricciones en los movimientos.	Filtro de condensación.	Determinación postura de la mano.
	Modelo de movimiento de cabeza de la Cascia, Isidoro y Sclaroff.	Cilindros	Alta precisión en el seguimiento de la cabeza.	El modelo tiene dificultades cuando se efectúan rotaciones rápidas y largas de la cabeza. Exige posición de la cabezza frontal. Alto coste computacional.	Template matching.	Determinación postura cabeza. Reconocimiento expresión facial.
	Modelo de movimiento de brazo de Luis Goncalvest. Bernardotl, Ursellaj y Peronat.	Conos Cilindricos	Alta precisión en el seguimiento de los brazos.	Alto coste computacional. Gran consumo de memoria a lo largo del tiempo.	Template matching.	Estimación de la postura del brazo.

Tabla 2.2: Resumen de estado del arte en Modelado de Humanos.

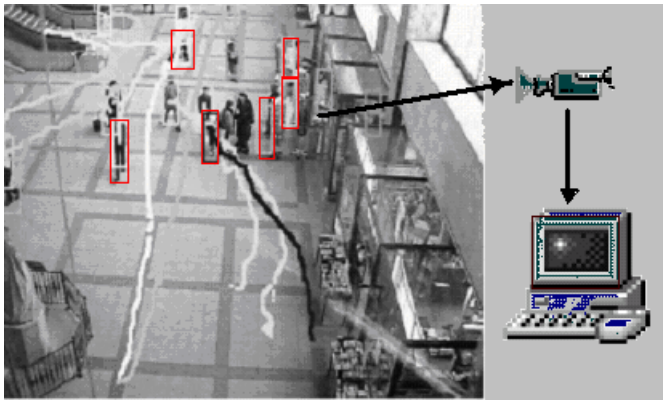


Figura 2.15: Frame de una secuencia de un lugar video-vigilado. Obtención de imágenes por la cámara de vigilancia y posterior procesado en el ordenador para su tratamiento.

monitorización de humanos, el cual trata de mejorar dos de los principales cuellos de botella en este tipo de sistemas de vigilancia, como son: la robustez y la velocidad. Para conseguirlo, se utiliza un modelo adaptativo basado en el algoritmo MeanShift, para obtener mejores soluciones a problemas habituales en en diversos escenarios, tales como la detección del color de la piel, el manejo de sombras y el análisis del movimiento. Csaba et al. (2005) describen un sistema de vigilancia también basado en meanshift para escenarios complejos con multitud de personas. En estos escenarios, el seguimiento de los individuos se complica debido a la aparición de múltiples oclusiones, la formación de grupos o la aparición de observaciones similares correspondientes a distintos individuos.

Con cámaras situadas a media distancia, los sistemas de vigilancia utilizan aproximaciones en 2D o en 3D: aproximaciones 2D cuando no se necesita una gran precisión en el seguimiento ni determinar la posición de las partes del cuerpo, y modelos en 3D para aplicaciones en entornos más exigentes, cuando se desea seguir a varios humanos que forman parte de un grupo, manejo de oclusiones y colisiones, permitir una gran variedad de movimientos, etc. Los modelos ocultos de markov han sido ampliamente empleados. En Lyons & Pelletier (2000), se utiliza un algoritmo llamado: “Nine–grid” para etiquetar características del cuerpo usando un modelo en 2D. Pfinder, Wren et al. (1997a), es un sistema en tiempo real que se utiliza se utiliza en sistemas de vigilancia reales. Utiliza un modelo estadístico multi-blob para el seguimiento de personas e interpretación de sus conductas.

En Lou et al. (2002), se consigue una descripción de la actividad realizada a partir del análisis de la secuencia de movimientos humanos utilizando Modelos de Markov (HMM). En Nair & Clark (2002) se utilizan estos modelos para detectar conductas anómalas mediante el uso de cámaras de seguridad colocadas en pasillos

dentro de una oficina. En Oliver et al. (2000) se describe un sistema real de visión para modelar y reconocer interacciones en tareas de vigilancia. Este sistema detecta interacciones entre personas y clasifica el tipo de interacción.

Leo et al. (2003) abordan el problema de detección de humanos en un entorno exterior utilizando una cámara estática para realizar tareas de vigilancia. Este sistema se prueba en el entorno de un parque, en el cual se resaltan los objetos de interés y se reconocen conductas relacionadas con robos o violencia. Se proponen técnicas para detectar personas mediante modelos dinámicos, tanto de forma individual, como en grupo. En otra publicación de los mismos autores Spagnolo et al. (2003), se plantea la estimación de posturas en el contexto de la vigilancia de un campo arqueológico. Siempre que un humano es detectado, sus posturas se clasifican mediante un módulo de estimación de dichas posturas. Luego los resultados son alimentados mediante un subsistema de HMM que identifica la actividad realizada. Las características seleccionadas para la estimación de las posturas se basan en histogramas verticales y horizontales de las formas obtenidas.

Otros trabajos, como Harwood & Davis (1998), emplean una combinación de análisis de la apariencia y seguimiento para detectar y seguir a múltiples personas, incluso para poder monitorizarlas en presencia de oclusiones y en exteriores. En el método descrito en Ju et al. (1996), se representa la posición relativa y el tamaño de las diferentes partes del cuerpo.

Cohen & Medioni (1999); Medioni et al. (2001) proponen utilizar eventos para modelar las acciones realizadas por humanos. El procesamiento de una secuencia de vídeo para caracterizar eventos de interés se basa en la detección, en cada frame, de los objetos implicados, y la integración temporal de esta frame para modelar comportamientos simples y complejos. Este alto nivel de descripción en una secuencia depende de la precisión en la detección y seguimiento de los objectos en movimiento y la relación de sus trayectorias. En la mayoría de las escenas hay un número significativo de objectos en movimiento y su análisis de sus trayectorias y la interacción con las características de la escena permiten clasificar y reconocer eventos de interés.

Thonnat & Rota (2000) tratan fundamentalmente de la interpretación de las imágenes de un video. La meta es detectar personas y analizar su conducta. El sistema propuesto está basado en un conocimiento a priori. Se basa en tres premisas : 1) se considera la cámara estática; 2) sólo se usará una única cámara monocular y 3) se harán varias restricciones a la hora de trabajar en tiempo real para así obtener buenos resultados en el mínimo tiempo de computación. En otras investigaciones relacionadas con la anterior, como Thonnat & Rota (1999), la meta es reconocer la conducta de un humano que está siendo seguido. El algoritmo consta de cuatro

fases: detección del movimiento, agrupamiento, fusión y supresión de agrupaciones innecesarias en el tratamiento. Usa información contextual en 3D en la fase de fusión y supresión.

En otros trabajos, como Ivanov et al. (1999), el sistema de vigilancia realiza un etiquetado de eventos e iteraciones en un entorno exterior. El sistema está diseñado para monitorizar actividades en un parking exterior, de modo que se controlen robos u otro tipo de ataques delictivos que pudieran tener lugar. Consta de tres componentes: un seguidor adaptativo, un generador de eventos, y un analizador. El sistema realiza la segmentación y el etiquetado de un video en el parking e identifica a las personas y su vehículo, estableciendo una asociación entre ambos.

Muchos sistemas de vigilancia trabajan con cámaras fijas, sin embargo, en cámaras de exterior o cuando la cámara está situada en una plataforma móvil es necesario utilizar algoritmos diferentes para realizar el seguimiento. Así, por ejemplo, Davis et al. (2000) utiliza modelos deformables junto con una variante del algoritmo de condensación para realizar el tracking de siluetas de humanos. Gavrila & Philomin (1999) presentan un sistema de vigilancia para ayuda a la conducción que detecta y distingue, en tiempo real, personas desde un vehículo en movimiento. Tiene algunas limitaciones relacionadas con el algoritmo de segmentación o con la posición en la que se encuentran los humanos, ya que el sistema no puede detectar personas muy cercanas a la cámara.

Debido al incremento de accidentes en las ciudades en los que están implicados peatones, también se han implementado sistemas para detectar acciones incorrectas (prohibidas). Por ejemplo, Tany (2004) propone usar imágenes tomadas con una cámara CCD situada a una distancia de aproximadamente 100m para detectar peatones en un paso de cebra y controlar así el tiempo de paso. Este sistema se ha implementado en algunas calles en Japón con un acierto del 99 %.

La cara es un rasgo muy característico y discriminante del ser humano. En **?** se propone un sistema de clasificación basado en una red neuronal multicapa, cuyas entradas serán muestras de fotografías faciales con diferentes variaciones de iluminación, postura y tiempo, con un volumen de muestras que simula un entorno real. La salida no es el reconocimiento del individuo como tal pero si la clase a la que pertenece. En Gutta et al. (1998)se describen sistemas de reconocimiento de caras para vigilancia. Se trata de identificar a una persona dentro una pequeña galería o grupo de imágenes previamente almacenadas, algunas de las cuales corresponden a potenciales intrusos. Los problemas asociados con la variabilidad de la imagen se mitigan usando una arquitectura conexionista, similar en su diseño a una mezcla de sistemas expertos. En Kruppa et al. (2003) se detecta inicialmente el contorno de la

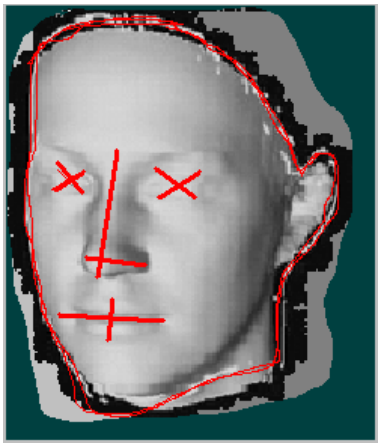


Figura 2.16: Selección del contorno del rostro y partes del mismo para reconocimiento en tareas de vigilancia

cara, así como del torso, para situar sobre un contexto local al detector (fig. 2.16). Por último, en Foresti et al. (2003) se define un sistema de detección de caras en secuencias de vídeo en color. El sistema utiliza una jerarquía de tres métodos para focalizar del análisis. El primer método localiza la cabeza del humano, enfocando todo el análisis a esta zona. Después, el segundo método detecta regiones de piel. Finalmente, se utiliza análisis de componentes principales para reducir la dimensión del conjunto de datos y detectar modelos de cara en dicho espacio.

En la última década se han desarrollado varios proyectos internacionales de vídeo-vigilancia. Entre los más relevantes, podemos citar los siguientes:

VSAM. Video Surveillance and Monitoring Collins et al. (2000). Este proyecto, desarrollado por Carnegie Mellon University y el Sarnoff Institute, propone un sistema para aplicaciones de vigilancia en ambientes complejos, con aglomeraciones, como en entornos urbanos o en el campo de batalla. El objetivo del sistema es proporcionar una cobertura continua de las personas y vehículos presentes en la zona vigilada a través de una amplia cantidad de sensores visuales situados en puntos estratégicos y facilitar la supervisión a un operador que puede seleccionar y monitorizar aquello que considere relevante y olvidarse del resto.

ADVISOR. Annotated Digital Video for Surveillance y Optimised Retrieval (Advisor) Siebel & Maybank (2004); Siebel et al. (2004) es un proyecto europeo para la vigilancia en estaciones de metro (fig.2.17). El sistema utiliza múltiples cámaras para realizar un seguimiento de las personas y detecta situaciones relacionadas con individuos y con grupos de indivi-

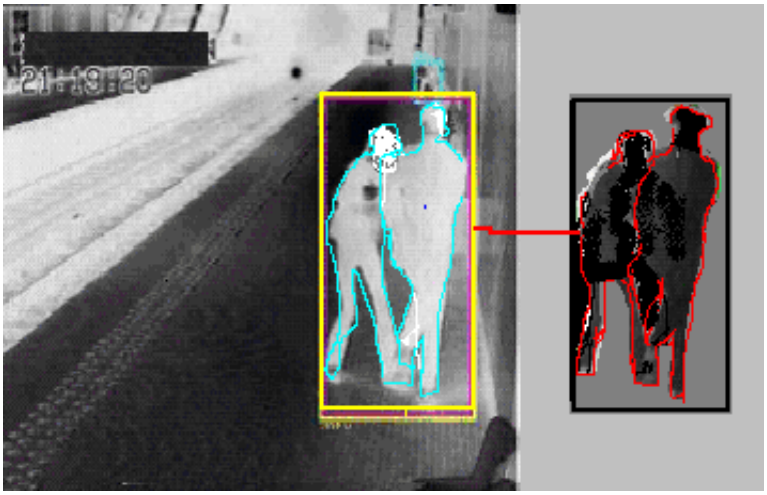


Figura 2.17: Aplicación de sistema de vigilancia a humanos en el metro. Extracción de la silueta de personas en el metro para su tratamiento.

duos, como violencia entre personas, vandalismo contra el mobiliario o personas saltando los controles de acceso. Además, detecta eventos relacionados con multitudes, como zonas masificadas, objetos abandonados, congestión en salidas y escaleras o movimiento de individuos en contra del flujo principal.

CSAIL. Computer Science and Artificial Intelligence Laboratory (CSAIL) Rahimi et al. (2004) es una parte del proyecto VSAM desarrollado por el grupo de Vision del MIT, el cual se centra en el desarrollo de técnicas optimizadas para el seguimiento de individuos dentro de una red de cámaras que cubren eficientemente una zona determinada que se pretende vigilar.

CAVIAR. Context Aware Vision Picture-based Active Recognition (CAVIAR: http://homepages.inf.ed. List et al. (2005) es un proyecto europeo cuyo objetivo principal era investigar en los temas centrales de la visión cognitiva, en concreto, mejorar los procesos de reconocimiento basados en imagen mediante el uso de sensores visuales con distintas arquitecturas organizativas (como los sensores foveados), descripción de la escena en distintos niveles semánticos, uso de conocimiento contextual y control espacio-temporal de la atención. El generó una arquitectura para el análisis cognitivo de streaming de video en la que era sencillo añadir nuevos módulos e incluso comparar fácilmente dos módulos que realizaran la misma función. La

arquitectura es distribuída, pero utiliza un controlador central que regula el funcionamiento de los módulos del sistema, por ejemplo, puede pedir a un módulo que encuentre más objetos o que genere menos características. Cada módulo obtiene datos de una o más fuentes y genera datos más información para realimentación, como información sobre cómo el módulo estimó que era la salida que produjo y qué necesitaría para mejorarla la próxima vez. El proyecto se centra en dos tipos de aplicaciones: 1) la vigilancia del centro de las ciudades para la detección de alcoholísmo, peleas, vandalismo, robos en tiendas, etc.; y 2) el análisis del comportamiento de los consumidores en una zona comercial para mejorar las ventas adaptándose a los potenciales clientes.

W4 W4 Haritaoglu et al. (2000) es un sistema de video-vigilancia en tiempo real para la detección y el seguimiento de varias personas y el seguimiento de sus actividades en un entorno al aire libre a partir de una sóla cámara (una cámara monocular y en escala de grises o una cámara de infrarrojos). W4 emplea una combinación de análisis de la forma y tracking para localizar a las personas y sus partes (cabeza, manos, pies, torso) y crear modelos de apariencia. El sistema puede segmentar varias personas que forman parte de un grupo y seguirlas de manera independiente. También puede determinar si las personas transportan objetos y segmentarlos, y reconocer eventos entre personas y objetos, tales como dejar un objeto, intercambiar bolsas o el robo de un objeto.

A modo de resumen, en la tabla 2.3 se presentan las publicaciones más citadas entre los años 2000 y 2010 ordenadas en función del número de citas, lo que nos da una idea del interés que cada una de ellas ha despertado entre la comunidad científica. Una revisión periódica de este tipo de información y de los resultados que proporciona su análisis debe favorecer el seguimiento de las tecnologías y extraer tendencias, así como detectar nuevas oportunidades tecnológicas. En la tabla 2.4 se presenta el número de publicaciones según la tecnología utilizada. En ésta, en primer lugar, se distinguen cómo las dos grandes tecnologías: una relacionada con la investigación en el reconocimiento facial y otra, en el de la biometría por voz.Estas dos primeras tecnologías (facial y voz) representan por referencias de términos relacionados más del 50 % de las publicaciones. Otro grupo por peso de publicación estaría formado por las tecnologías basadas en huella dactilar, firma escrita, y análisis gestual. El último grupo englobaría a tecnologías más incipientes, con un interés

Autor	Título	Fuente	Año publicación	citas
Phillips, PJ; Moon, H; Rizvi, SA; Rauss, PJ	The FERET evaluation methodology for face-recognition algorithms	IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE	2000	1000
Zhao, W; Chellappa, R; Phillips, PJ; Rosenfeld, A	Face recognition: A literature survey	ACM COMPUTING SURVEYS	2003	971
Yang, MH; Kriegman, DJ; Ahuja, N	Detecting faces in images: A survey	IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE	2002	884
Georghiades, AS; Belhumeur, PN; Kriegman, DJ	From few to many: Illumination cone models for face recognition under variable lighting and pose	IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE	2001	583
Martinez, AM; Kak, AC	PCA versus LDA	IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE	2001	477
Hsu, RL; Abdel-Mottaleb, M; Jain, AK	Face detection in color images	IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE	2002	465
Plamondon, R; Srihari, SN	On-line and off-line handwriting recognition: A comprehensive survey	IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE	2000	438

Tabla 2.3: Publicaciones científicas más relevantes en vigilancia por número de citas. Boletín de vigilancia tecnológica que se enmarca dentro de la cátedra de Propiedad Industrial e Intelectual de Clarke, Modet & Co. con la Universidad Politécnica de Madrid (2011).

Tecnología Biométrica	Nº Publicaciones
Reconocimiento facial	3118
Patrón de la voz	2375
Huellas digitales	867
Análisis gestual	744
Firma y escritura manuscrita	672
Cadencia de paso	530
Análisis del iris	411
Rayas de la mano	237
Geometría de la mano	94
ADN	84
Análisis de la retina	76
Reconocimiento vascular	33
Dinámica de tecleo	33
Termografía general	13
Olor corporal	11
Análisis de la córnea	9

Tabla 2.4: Número de publicaciones relacionadas con la tecnología biométrica entre (2000-2010). Boletín de vigilancia tecnológica que se enmarca dentro de la cátedra de Propiedad Industrial e Intelectual de Clarke, Modet & Co. con la Universidad Politécnica de Madrid (2011).

investigador todavía menos significativo, como aquellas basadas en la geometría de la mano, la termografía, el olor corporal, o los análisis de retina o córnea. Esto da una muestra del interés por el análisis de situaciones relacionadas con humanos. En términos económicos, el mercado biométrico crece cada año y fuerza una mayor demanda de la industria implicada, de desarrolladores de soluciones, de investigadores y de usuarios finales.

2.2. Reconocimiento de actividades

Tal como se describe en Martinez-Tomas et al. (2008), las actividades se pueden modelar como eventos complejos, los cuales son definidos por composición espacio-temporal de eventos más simples, éstos a su vez por otros eventos aún más simples y así sucesivamente formando una jerarquía de eventos hasta enlazar con eventos primitivos, los cuales se determinan a partir de cambios de estado de atributos visuales. En Folgado et al. (2011)nos centramos en el estudio de un modelo de humanos y su aplicación se enmarca dentro del nivel de objetos de la jerarquía de los diferentes niveles de descripción de la escena, siendo un nivel intermedio entre el de blob y actividad.

Como se ha podido comprobar en las secciones anteriores, este es un campo muy

activo actualmente, tal como se ve reflejado en el gran número de trabajos publicados. En general, las soluciones son fuertemente dependientes de los objetivos y los campos de aplicación son múltiples (vigilancia, estudios médicos y de rehabilitación, robótica, indexación de video y videojuegos, etc). En Poppe (2010) se describen los principales enfoques existentes para etiquetar acciones humanas sobre secuencias de video.

. Muchos trabajos tratan de aprender y reconocer actividades de forma indirecta, simplemente observando el movimiento de los objetos y buscando la correlación con las actividades objetivo. Para ello, no es necesario niconocer la identidad del sujeto, ni una descripción detallada de la actividad. Así Stauffer & Grimson (2000); Efros et al. (2003) tratan de reconocer acciones simples de personas (correr, pegar patadas o saltar verticalmente, etc), en imágenes de video de baja calidad donde las personas tienen un altura de aproximadamente 30 píxeles, para lo cual, usan un conjunto de características que se basan en el flujo óptico difuminado. Otros como Robertson & Reid (2005), tratan de reconocer acciones construyendo un sistema jerárquico que se basa en razonar con redes neuronales y modelos de Markov para un análisis a nivel alto. Este reconocimiento de actividades no sólo tiene sentido en entornos comunes, como edificios, parques, calles o instalaciones de transportes, si no en escenarios de cualquier tipo, por ejemplo, en Eng et al. (2003) se presenta un sistema de videovigilancia de una piscina a partir de información como la velocidad, postura, índice de inmersión y un índice de actividad de los objetos.

Otro gran número de publicaciones trabajan en el concepto de espacio-tiempo. Una de las principales aportaciones Rittscher et al. (2002) usa patrones espacio temporales a partir del volumen, donde los movimientos de las articulaciones pueden asociarse con modelos de trayectoria y utilizar éstos para detectar acciones. Bobick (1997), proponen una representación que se basa en imágenes de energía del movimiento e imágenes de historia del movimiento. Yi Yi et al. (2004) presenta la idea de un mapa de relación de cambio de píxeles. Sin embargo, posteriores procesamientos se basan en histogramas de movimiento. Ozer & Wolf (2002), abordan el seguimiento, la estimación de la postura y el reconocimiento de una forma integrada, mientras que en Gao et al. (2004), se lleva a cabo un análisis de las actividades, combinando segmentación con seguimiento. Vecchio et al. (2003), usan técnicas de un sistema dinámico para logar la segmentación y clasificación. Yu & Yang (2005) utilizan redes neuronales para encontrar primitivas en los que aplican mapas auto-organizados (SOM) que agrupan las imágenes de aprendizaje basándose en información de la forma. Después del aprendizaje SOM se genera una etiqueta para cada imagen de entrada que convierte una secuencia de imágenes en una secuencia de etiquetas.

Existen otros métodos, como en Meeds et al. (2008), que usan modelos gráficos probabilistas para inferir modelos para la estimación de la postura, oclusiones, etc, sin necesidad de conocimiento previo del tipo de movimiento.

Por otra parte, algunas aplicaciones de video-vigilancia se centran explícitamente en el reconocimiento de actividades realizadas por una persona o actividades que implican interacción Sato & Aggarwal (2001). En este caso, se distinguen dos posibilidades: que se busque una descripción de la persona en su conjunto, sin tener en cuenta sus partes, como por ejemplo, , información del género, de la identidad o de acciones simples como caminar o correr Cheng et al. (2002), o que se busque una descripción más detallada, en la que se intenta modelar primitivas de acción simples que por composición permiten modelar acciones más complejas Leo et al. (2004)(coger una mochila, saluda, abre una puerta, etc). Estos eventos primitivos relacionados con humanos se pueden utilizar para la descripción de actividades más complejas en multitud de tareas de vigilancia (vigilancia de equipajes en aeropuertos, vigilancia de comportamientos de clientes en bancos, vigilancia de pacientes en hospitales, etc) . Así, algunos autores tratan el reconocimiento de acciones basándose en la dinámica y la disposición de las partes individuales del cuerpo, como en L.Wang et al. (2003), que presenta un trabajo donde se extraen los contornos y la postura. Otros trabajos reconocen el mismo tipo de acciones pero se fundamentan en modelos de Markov (HMM) Elgammal et al. (2003); Luo et al. (2003). Parameswaran & Chellappa (2003) consideran el problema de reconocimiento de acciones mediante curvas invariantes al punto de vista.

En Gonzalez et al. (2002) se usa un modelo de distribución de puntos para modelar la dinámica de los ángulos de las articulaciones de un modelo humano sencillo. Para cada acción que se desea reconocer, se utiliza la configuración de parámetros obtenida para un conjunto de individuos realizando dicha acción para obtener un sistema que la reconoce mediante aprendizaje supervisado.

En Calinon et al. (2005) se presenta una solución basada en HMM para aprender características de movimientos repetitivos. Sugieren el uso de HMM para sintetizar las trayectorias de las articulaciones de un robot, para cada articulación se usa un HMM. Fanti et al. (2005) proponen un modelo de conocimiento en el que se combina información de distinta naturaleza (posiciones, velocidades y apariencia) con variables globales tales como traslación, escala o punto de vista para mejorar el rendimiento del sistema a la hora de reconocer distintos tipos de movimiento humano (andar, correr, pedalear, etc.). Otras acciones, tales como llevar bolsas de diferentes pesos se puede realizar basándose en una descripción de las trayectorias en base a descripciones de la pose y su evolución temporal y del esfuerzo Davis &

Gao (2003).

Lu & Ferrier (2004) abordan el problema desde un punto de vista teórico, su objetivo es segmentar y clasificar eventos a partir de otros más simples. En Rao et al. (2002) proponen una representación invariante de las acciones basándose en intervalos e instantes dinámicos para describir eventos complejos como recoger bolsas, cerrar y abrir puertas, etc. Utilizando PCA's en Cho et al. (2009), para analizar la enfermedad de Parkinson en pacientes basándose en su forma de "andar".

También es posible clasificar los trabajos por la dimensionalidad de los modelos utilizados (2D o 3D). Como ejemplos 2D, podemos citar Roberts et al. (2004), Zhang et al. (2004), Sidenbladh et al. (2000), Boulay et al. (2003), Bobick & Davis (2001) yCohen & Li (2003). Otras propuestas describen el movimiento de todo el cuerpo humano desarrollando varios sistemas que permiten realizar un seguimiento en 2-dimensiones (2D) y trasladarlo a 3-dimensiones (3D). Como ejemplos 3D podemos citar Heisele & Wöhler (1998); Zhao & Thorpe (2000); Sminchisescu et al. (2004); Sminchisescu & B.Triggs (2003); Ong & Gong (1999); Lan & Huttenlocher (2005); Ramanan & Forsyth (2003); Ikizle & Forsyth (2008). Trabajos como Ikizle & Forsyth (2008) han mostrado que los movimientos atómicos de las diferentes partes del cuerpo en 3D, cuando son combinados con modelos ocultos de Markov (HMMs), se pueden utilizar para inferir composiciones de movimientos complejas. También existen algunas aplicaciones híbridas que mezclan representaciones 2D y 3D Boulay et al. (2005); Pan (2000).

En la tabla 2.5 se muestran algunos enfoques utilizados para el reconocimiento de actividades, así como las actividades reconocidas.

El uso de modelos de humanos para la interpretación de secuencias de video forma parte del estado del arte en visión artificial y puede ser contrastado en la amplia bibliografía existente sobre el tema, tal y como se ha mostrado. Actualmente, se están utilizando modelos de humanos para la descripción de comportamientos de los humanos presentes en la escena, pero los sistemas resultantes son sistemas no robustos y están basados en unas suposiciones muy restrictivas que simplifican las primeras fases del procesamiento de las imágenes (segmentación, reconocimiento y seguimiento), tal y como, se ha analizado en la bibliografía. Se podría decir que la creación de un módulo de seguimiento de humanos robusto es primordial para la generación de sistemas de visión automáticos para la interpretación de escenas. Los fines de estos sistemas son muy variados: vigilancia, interfaces de usuario, anotación de secuencias de imágenes para indexación y búsqueda en bases de datos, etc. La investigación en reconocimiento de actividades persigue avanzar en el uso de modelos de humanos en visión artificial y en el desarrollo de sistemas de seguimiento de

Referencia	Tipo de Enfoque	Actividades reconocidas
K. Sato and J.K. Aggarwal	Espacio-Tiempo (Trayectorias) con el método SVM.	Interación entre dos personas. Abrazo, empujar, dar la mano, etc.
C. Rao, A. Yilmaz, and M. Shah	Espacio-Tiempo (Trayectorias) con el método Template Matching.	Abrir, cerrar armarios, puertas, recoger objectos, bolsas, etc.
Bobick, A.F., Davis, J.W	Espacio-Tiempo (volumen) con el método Template Matching.	Reconocimiento de diferentes ejercicios aeróbicos.
Fangxiang Cheng, William J. Christmas and Josef Kittler	Utiliza métodos de clasificación Bayesianos.	Distingue entre andar o correr, a partir de un video.
Ch. Cho, W. Chao, W.H. Lin, Y.Chen.	Técnicas de Análisis de Componentes Principales (PCA)	Detecta la enfermedad del Parkinson por la manera de andar.
H.L. Eng, K.A. Toh, A.H. Kam, J. Wang, and W.Y. Yau	Modelos de Markov	Nadando normal, con angustia, sacando el pie, etc.

Tabla 2.5: Enfoques utilizados para reconocimiento de actividades.

humanos robusto, un problema todavía sin resolver.

La mayoría de las aproximaciones para el análisis de la actividad se basan en la definición de modelos específicos para una determinada actividad en un determinado dominio de aplicación, siendo altamente dependientes de los resultados obtenidos durante la etapa de seguimiento. El problema fundamental es el enorme salto semántico que hay entre el nivel físico de las señales y el nivel de conocimiento. Esto lleva a que los objetivos de los sistemas sean poco ambiciosos, pues la probabilidad de fallo es muy alta en cuanto tratan situaciones o actividades complejas.

2.3. Reconocimiento de personas por la forma de andar (Gait)

El concepto de análisis biométrico de la forma de andar, en inglés "gait analysis", aparece hacia 1994 Niyogi & Adelson (1994). Este análisis puede tener dos objetivos: 1) identificar la dinámica del movimiento, lo que puede ser aplicado en medicina o deportes, por ejemplo; y 2) identificar a individuos concretos, lo que puede ser utilizado en vigilancia. La utilidad principal dentro del campo de la vigilancia es que es un sistema no invasivo y que permite adelantar la identificación a cuando el individuo está aún alejado, a una distancia de seguridad.

El libro de Nixon et al. (2006) incluye las principales técnicas, sistemas, bases de

datos y trabajos dentro de la identificación humana basada en el gait. Más recientemente, Sarkar et al. (2005) tratan de caracterizar las propiedades del gait mediante un conjunto de doce experimentos que examinan la influencia de distintas variables, denominadas covariables. Cada experimento consta de las llamadas definiciones de galeria (watch-list) y de los datos de entrada que difieren con respecto a una o más covariables. Estas covariables son: el tipo de superficie (hierba, cemento), llevar maletín (si, no), cámara (izquierda, derecha), tipo de zapato (A, B) y tiempo (Mes Mayo o Noviembre). Con los experimentos se examina el efecto sobre el rendimiento de diferentes ángulos respecto a la cámara, cambios en la superficie, etc. La base de datos utilizada para realizar dichos experimentos se denomina HumanID Gait Challenge Problem.

Boulgouris & Chi: (2007) se basan en el análisis por separado de diferentes componentes del cuerpo que son identificables en la silueta del humano (cabeza, torso, brazos, piernas y muslos) para investigar la influencia de cada componente del cuerpoen el sistema de reconocimiento. Para ello, se etiquetan manualmente las siluetas y se estudia cada componente de manera independiente.

Varios trabajos analizan la identidad de un individuo examinando su patrón de marcha al caminar. Por ejemplo, Boyd & Little (2005) consideran el "Gait" como una combinación de varios movimientos. Los movimientos son coordinados en el sentido de que ocurren siguiento un patrón temporal que se repite de manera cíclica con los pasos del humano al caminar. El análisis de este patrón permite identificar al individuo. En Yam et al. (2002) se investiga la relación entre los movimientos de caminar y correr y construye un único algoritmo basado en el análisis de Fourier del movimiento de los muslos y de los gemelo de las piernas para reconocer a las personas tanto caminando como corriendo.

Utilizando como características de entrada las variaciones de señales de distancia extraídas de la silueta del individuo, en Wang et al. (2003) se propone un sistema de reconocimiento por el gait con poca carga computacional basado en técnicas de reconocimiento de patrones basadas en medidas de similitud respecto a patrones predefinidos, mientras que en Kale et al. (2002) se utilizan modelos Markov para definir el modelo.

En Bouchrika & Nixon (2007); Goffredo et al. (2008) se utiliza la evolución de la silueta para extraer parámetros relacionados con la forma del cuerpo (ancho, alto, longitud de las extremidades, etc) o con el movimiento (longitud del paso, tiempo periódico, velocidad, ángulos entre extremidades, etc).

En resumen, podríamos decir que reconocer una persona por la forma de caminar es una tarea muy compleja pues, tal como aseguran estudios recientes Bouchrika &

Nixon (2008); Guo & Nixon (2009), la forma de caminar se ve claramente afectada cuando se modifica la apariencia del individuo, sobre todo si se modifica la silueta por motivos tales como llevar puesto un abrigo o un sombrero, tirar de una maleta, etc.

2.4. Métodos de aprendizaje utilizados en la tarea de reconocimiento de actividades y eventos

La construcción de modelos para la descripción de los estados y eventos primitivos relacionados con la descripción del comportamiento humano y de las actividades de interés para tareas de vigilancia o monitorización es una tarea sumamente compleja. En las últimas décadas, el gran desarrollo y abaratamiento de las tecnologías de adquisición de imágenes y de comunicaciones han facilitado la obtención de cantidades masivas de datos relacionados con las personas y sus acciones. Esto ha permitido utilizarlos para describir situaciones de interés mediante ejemplos representativos y utilizar técnicas de aprendizaje automático supervisado para el modelado de las situaciones de interés.

El reconocimiento de acciones y eventos en vigilancia se trata principalmente como un problema de clasificación, donde el objetivo es seleccionar la clase más cercana entre un conjunto de clases predefinidas. El comportamiento de un humano en una escena es claramente dinámico, ya que sus acciones dependen tanto de estímulos exteriores como de su historia anterior, lo que hace necesario utilizar modelos dinámicos para su descripción computacional. Sin embargo, muchos eventos simples se pueden describir mediante modelos estáticos, en los que la variable tiempo se hace constante.

Teniendo en cuenta todo lo anterior, en esta sección comentaremos los métodos de clasificación mediante aprendizaje automático supervisado, tanto estáticos como dinámicos, que más comúnmente se han utilizado en aplicaciones de vigilancia.

2.4.1. Clasificadores basados en modelos estáticos

Los métodos de clasificación mediante aprendizaje automático supervisado basados en modelos estáticos más comúnmente utilizados en aplicaciones de vigilancia han sido los árboles de decisión, las máquinas de vectores soporte, las redes neuronales y más recientemente los multiclasificadores. Ver capítulo 7 para detalles de uso de herramientas de clasificación estática.

Los árboles de decisión construyen un modelo jerárquico en el cual se traza un mapa de atributos y nodos, donde cada rama desde la raíz a un nodo de hoja es

una regla de clasificación. Los árboles de decisión son de los métodos de aprendizaje más usados, la justificación de este uso es debida a que son fáciles de comprender e interpretar y dan una idea bastante aproximada de las variables más relevantes Quinlan (1993); Jatoba et al. (2008); Maurer et al. (2006). Dentro del campo de reconocimiento de actividades, en Bao & Intille (2004) se propone un sistema que reconoce mediante clasificadores basados en arboles de decisión hasta 20 actividades, tales como fregar, aspirar, ver la televisión, trabajar en el PC, conducir, etc. Este trabajo sugiere además que con pequeños acelerómetros inalámbricos, colocados en el muslo de un individuo y la muñeca, se pueden detectar algunas de las actividades cotidianas comunes en ambientes cotidianos con rapidez y de manera sencilla. Sus resultados en la detección de este tipo de actividades está en torno al 84 %. En otro trabajo Ermes et al. (2008), se reconocen las siguientes actividades: tumbado, sentado, de pie, andando, corriendo y haciendo ciclismo, etc. El análisis se realiza "offline" con los datos recogidos por diferentes sensores y enviados a una PDA (Personal digital assistant), donde un algoritmo propio analiza los datos sensoriales para reconocer la actividad y proporcionar una tasa de acierto de aproximadamente el 86 %. La PDA calcula las características de la señal que se utilizan para el reconocimiento de actividades provenientes de las señales de aceleración, realizando posteriormente la clasificación de actividades. Las características de la señal consideradas son: la media, la varianza, la frecuencia con mayor pico y la entropía. Con la media de la señal se detecta la posición del cuerpo y con la varianza la intensidad de la actividad.

Las máquinas de vectores de soporte (SVM) Cortes & Vapnik (1995); Jhuang et al. (2007); Laptev et al. (2007); Schüldt et al. (2004) y las redes neuronales Gallant (1990); Randell & Muller (2000) también han sido ampliamente utilizadas en reconocimiento de actividades He & Jin (2008); He et al. (2008); He & Jin (2009), aunque sus reglas, a diferencia de los árboles de decisión, no son tan compresibles, se llega a obtener una mayor tasa de acierto. Las SVM's dependen de las funciones denominadas kernel, las cuales proyectan todas las instancias a un espacio de dimensión superior con el fin de encontrar las frontera de decisión entre las clases. Las SVM's presentan una enorme robustez y desde su introducción en los años setenta han supuesto un gran avance en las técnicas de aprendizaje a partir de muestras Cortes & Vapnik (1995). Estas técnicas han sido utilizadas en muchas aplicaciones, tales como, reconocimiento de voz Ramírez et al. (2006), tareas de diagnóstico de imágenes Fung & Stoeckel (2007) o en clasificación de texturas Kim et al. (2002).

Las redes neuronales replican el comportamiento biológico de las neuronas en el cerebro humano, propagando las señales de activación y codificación a través de

los enlaces de la red. El uso de estos sistemas se debe más al éxito obtenido en aplicaciones reales (reconocimiento de patrones, predicción, optimización, etc) que a la semejanza con el modelo biológico. El perceptrón multicapa, por su simplicidad, es uno de los tipos de redes neuronales más utilizados. Algunas aplicaciones en las que han sido utilizados son el reconocimiento de caras Hancock et al. (1998); Lawrence et al. (1997), detección de actividades, tales como, andar, correr y caminar Randell & Muller (2000), el reconocimiento de personas por la forma de andar Kusakunniran et al. (2010) o la monitorización de la actividad física Jafari et al. (2007).

Los métodos multiclasificador se han utilizado en campos tan variados como las finanzas Leigh et al. (2002), la química Merkwirth et al. (2004), la medicina Mangiameli et al. (2004), el tratamiento de imágenes Lin et al. (2006), el tratamiento de información en procesos de fabricación Maimon & Rokac (2004) o la geografía Bruzzone et al. (2004) obteniéndose muy buenos resultados. Así, en Minnen et al. (2007) se proporciona una solución integrada que incluye la captura de datos y el reconocimiento automático de la actividad de un soldado. El sistema presenta una interfaz multimedia que combina la búsqueda de datos y la exploración. El componente de reconocimiento analiza las lecturas de seis acelerómetros repartidos por el cuerpo para identificar la actividad (correr, andar, etc). Las actividades son modeladas mediante métodos de "boosting", que permiten una selección eficiente de las características más relevantes. Siguiendo con este tipo de métodos, en Fathi & Mori (2008) se presenta un método para el reconocimiento de la acción humana en el marco de la vídeo-vigilancia que ha dado lugar a una base de datos con patrones de movimiento generada mediante el algoritmo AdaBoost a partir de características basadas en información de flujo óptico. Este método se ha aplicado posteriormente a bases de datos de futbol, ballet o al movimiento humano en general. En Laptev & Pérez (2007) se aborda el reconocimiento y la localización de las acciones humanas sobre diferentes episodios de películas con una variación de acciones en cuanto al tipo de movimiento, ángulos de visión, escenarios, etc, utilizando clasificadores de "boosting". Nowozin et al. (2007) proponen utilizar el clasificador LPBoost combinado con un número pequeño de funciones de decisión con el fin de detectar la presencia de algún tipo de patrón discriminativo dentro del conjunto de imágenes. Se utiliza el clasificador LPBoost para aprender simultáneamente una función de clasificación y la selección de características en el espacio de todas las secuencias elegidas. Reconoce actividades como boxear, aplaudir, agitar la mano, correr o caminar. En Smith et al. (2005), tienen la capacidad de mejorar clasificadores débiles permitiendo usar un histórico de las acciones realizadas para evaluar el actual frame. Se han utilizado algoritmos de boosting para determinar acciones tales como hablar

por teléfono, coger vasos, ponerse gafas, cascos, frotarse los ojos, etc.

2.4.2. Clasificadores basados en modelos dinámicos

Para el modelado del dinamismo inherente a los humanos se utilizan habitualmente métodos gráficos probabilistas, en concreto, los más utilizados son los modelos de Markov, ver capítulo 8. Estos modelos permiten estudiar la evolución temporal de cualquier proceso que cumpla la *propiedad de Markov:* el estado futuro de un sistema depende solo del estado en que se encuentre en el presente, pero no de su historia pasada. Los Modelos Ocultos de Markov (HMM) se utilizan para modelar sistemas que tienen estados inobservables, en este caso los estados tienen asociadas distribuciones de probabilidades sobre las posibles salidas. Debido a su naturaleza probabilística es posible diseñar un procedimiento para encontrar una estimación de los parámetros del modelo. El procedimiento consiste en encontrar los estimadores de máxima verosimilitud de los parámetros para una secuencia o secuencias de salida. Máximos locales de la verosimilitud se pueden encontrar de manera eficiente mediante el algoritmo de Baum-Welch, Baum et al. (1970); Welch (2003) y sus alternativas Davis & Lovell (2003); Matsuyama (2003, 2011).

El uso inicial de estos métodos fue en aplicaciones de reconocimiento de voz Schmidbauer (1989); Deng & Erler (1991); Deng & Sun (1994). Las aplicaciones de los modelos de Markov en el campo del reconocimiento de actividades han sido múltiples Zhu & Sheng (2009); Pham & Abdelzaher (2008); Vinh et al. (2011) puesto que permiten modelar de manera natural el movimiento humano, que es dinámico. Como ejemplos de aplicaciones, podemos citar el reconocimiento de gestos Starner & Pentland (1995), reconocimiento de acciones Green & Guan (2004) o para reconocimiento de actividades específicas dentro de un campo concreto, como es el caso del Tai Chi Brand et al. (1997). En este caso, se utilizan los modelos de Markov para clasificar tres tipos de movimiento diferentes de esta actividad, codificando la coordinación de las distintas partes del cuerpo de manera independiente.

Otros trabajos utilizan variaciones de los modelos de Markov, los llamados factoriales y jerárquicos. En los Modelos de Markov factoriales los estados del modelo original se factorizan en un número variable dando lugar a una topología distribuida Ghahramani & Jordan (1997), donde la probabilidad conjunta para la secuencia de estados y de observaciones puede ser factorizada en diferentes HMM. Estos se han orientado fundamentalmente en el reconocimiento de personas caminando o de gestos Chen et al. (2009). Los denominados modelos de Markov jerárquicos Fine et al. (1998), son aquellos en los que se asume una cierta estructura en el conjunto o en la sucesión de estados ocultos, lo que los hace adecuados para el reconocimiento de acciones formadas por la combinación de diferentes actividades.

Capítulo 3

Descripción del modelo de humanos BB6-HM

El objetivo final de un sistema de visión artificial es la descripción de la escena orientada a una determinada tarea. En esta tesis nos vamos a centrar en la tarea de la descripción de la actividad humana a partir de imágenes de vídeo. Una forma de describir la actividad humana es mediante composiciones espacio-temporales de actividades más simples, que a su vez se pueden descomponer en secuencias de actividades aún más simple, y así hasta llegar a secuencias de movimientos o eventos que no admiten descomposición, los movimientos o eventos primitivos. El nivel de granularidad a usar, o sea, qué se considera primitivo, dependerá del contexto de uso: en un contexto neurofisiológico habría que llegar a un nivel neuronal, en uno de vigilancia corresponden a conceptos visibles en la secuencia y que se pueden modelar mediante sistemas estáticos o dinámicos, como andar, desplazarse 20 metros hacia el sur, pararse, soltar un objeto, etc.

En este capítulo vamos a introducir las características de un modelo de representación de humanos que permita analizar secuencias de vídeo para monitorizar, en tiempo real y con una carga computacional mínima, una gran cantidad de estos eventos primitivos relacionados con las actividades realizadas por humanos. Estos eventos primitivos pueden utilizarse a modo de librería para la descripción de actividades más complejas en multitud de tareas de vigilancia: vigilancia de equipajes en aeropuertos, vigilancia de comportamientos de clientes en bancos, supervisión de pacientes en hospitales, monitorización de personas dependientes en el hogar, etc.

La manera más habitual de presentar el cuerpo humano, como se ha visto en el capítulo anterior, es mediante un conjunto de segmentos (barras o volúmenes) que están unidos entre sí a través de articulaciones. Esta representación se basa en la observación de que el movimiento humano es esencialmente el movimiento del

esqueleto humano asistido por los músculos adyacentes. La geometría de cada uno de los segmentos varía entre autores y es dependiente de la representación. En nuestra propuesta, utilizaremos un enfoque distinto: el modelo propuesto, denominado BB6-HM (*block-based-six human model*), consiste en seis bloques de la misma altura obtenidos al dividir verticalmente la silueta de un humano erguido. Esta división se inspira en las reglas de proporcionalidad utilizadas en Bellas Artes a la hora de representar un boceto de humano. Los bloques vienen caracterizados por un conjunto de parámetros que formarán un vector de características y que podremos utilizar posteriormente para reconocer distintas situaciones o actividades. Esta propuesta tiene la gran ventaja de descomponer el análisis del humano en partes, de modo que podemos obtener información (oclusiones, acarrear objetos, dejarlos, cogerlos, agacharse, levantarse, ...) centrando el análisis en partes concretas del humano y "olvidándonos" del resto.

La estructura del sistema global se define en la figura 3.1: 1) se partirá de una secuencia de imágenes previamente segmentadas para obtener la región de la imagen (blob) asociada a un humano; 2) una vez segmentada la imagen, se procederá a la definición del modelo de descripción del humano BB6-HM asociado a esa secuencia segmentada; 3) a partir de éste, se utilizarán diferentes modelos específicos para describir cada uno de los eventos primitivos que se pueden detectar a partir del modelo BB6-HM y que permiten describir situaciones tales como saltar, caminar, agacharse, etc. A partir de estos eventos primitivos se pueden definir eventos más complejos y a partir de éstos otros más complejos aún, de manera que se pueda conseguir la descripción y el reconocimiento de actividades humanas complejas. En esta tesis nos centraremos en el reconocimiento de situaciones relacionadas con un único humano, tales como andar, levantarse, coger un objeto, etc.), y en el reconocimiento de personas por su forma de andar (gait).

En la primera sección de este capítulo (3.2) se describirá el modelo BB6-HM. A continuación, en la segunda sección (3.2) se detallará el conjunto de parámetros que lo caracterizan. Finalmente, en la tercera sección (3.3) se analizará la información que se puede extraer de una determinada situación a partir de los parámetros asociados al modelo, en concreto, información sobre movimiento, periodicidad, posturas y situaciones excepcionales como oclusiones y acarreo de objetos.

3.1. Contexto de la tesis propuesta

Esta tesis, como se comentó al principio del capítulo 2, se encuadra dentro del contexto de la vigilancia visual (fig. 3.2), cuyo objetivo es monitorizar el entorno,

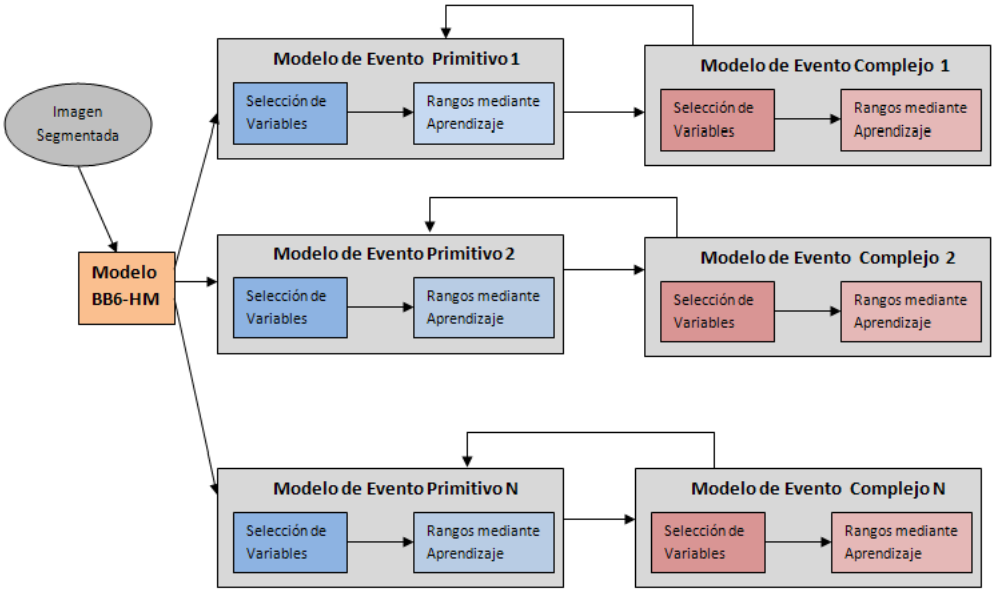


Figura 3.1: Esquema general del sistema de reconocimiento de situaciones utilizando el modelo de humanos BB6-HM. Una vez segmentada la secuencia se procederá a la descripción en base al modelo BB6-HM y se crearán diferentes modelos específicos por cada uno de los eventos y actividades detectadas

diagnosticar las situaciones (relaciones espacio-temporales entre distintos objetos de interés en una secuencia de imágenes) y generar las acciones pertinentes, en colaboración con los agentes humanos, ante situaciones de alerta. En nuestro caso, el marco conceptual de partida lo proporciona Martinez-Tomas et al. (2008), donde se define una arquitectura de niveles de descripción: nivel de imagen, nivel de blob, nivel de objeto y nivel de actividad. En cada nivel es necesario definir las entidades y relaciones de su ontología, así como los lenguajes de comunicación entre niveles. El trabajo de esta tesis se centra en el nivel de objetos, en concreto en objetos de tipo "humano" tal como se muestra en la figura 3.3. Nuestro objetivo con el modelo BB6-HM será caracterizar el dinamismo de un humano presente en la escena y detectar eventos y situaciones primitivas realizadas por éste a partir de una segmentación previa realizada en el nivel de blobs. Estos eventos y situaciones primitivos serán utilizados, en el nivel de actividades, para describir comportamientos más complejos y más cercanos a la tarea de vigilancia.

El modelo propuesto en esta tesis ha sido desarrollado dentro de los proyectos de investigación TIN2004-07661-C02-01, TIN2007-67586-C02-01 y UNED-2006-TrackingSystem, dando lugar a varias publicaciones Folgado et al. (2011); Rincón et al. (2007); Folgado et al. (2009); Martinez-Tomas et al. (2008) y una patente Rincón et al. (2006).

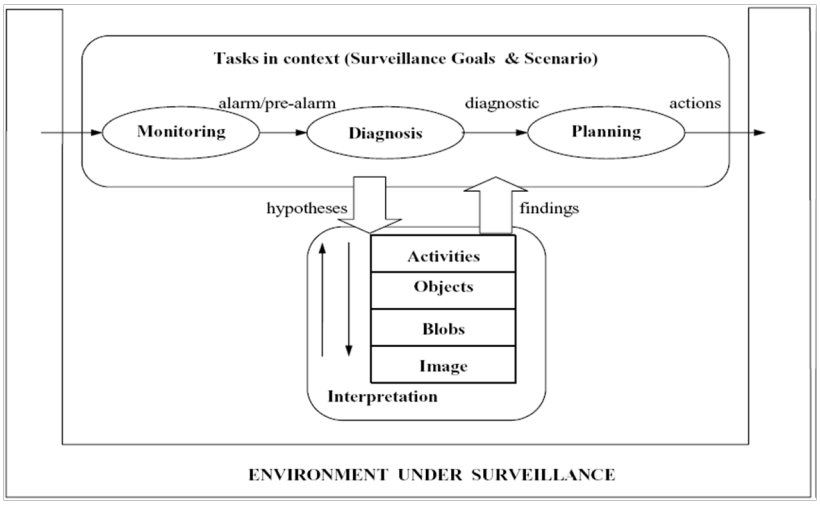


Figura 3.2: Tareas dentro del contexto de la vídeo-vigilancia (tomada de Martinez-Tomas et al. (2008)).

3.2. Descripción del modelo BB6-HM

El modelo propuesto está inspirado en las reglas de proporcionalidad utilizadas en Bellas Artes a la hora de estructurar el lienzo con el primer esbozo de una figura humana, las cuales provienen, seguramente, del canon de proporciones entre distintas partes del cuerpo desarrollado por Vitruvio e inmortalizado por Leonardo Da Vinci en su famoso "Hombre de Vitruvio" (fig. 3.4).

El modelo BB6-HM consiste en dividir verticalmente el blob correspondiente a un humano, el cual se ha segmentado previamente, en seis regiones de la misma altura (ver bloques B_1 , ..., B_6 en la fig. 3.5). Cada una de estas regiones puede ser delimitada por un rectángulo de selección que llamaremos "bloque". Cada una de estas regiones corresponde a una zona determinada del cuerpo humano cuando la persona está de pie, erguida y con los brazos hacia abajo:la cabeza está en B_1, los hombros en B_2, los codos en B_3, las manos y cadera están en B_4, las rodillas en B_5 y los pies en B_6.

Este modelo permite trabajar con vistas laterales y frontales, queda por tanto descartado el análisis de la vista cenital. En una vista frontal, el humano está mirando a la cámara o está de espaldas a ella. En una vista lateral, el humano está situado lateralmente respecto a la cámara. Estos puntos de vista se pueden medir por el ángulo de visión: una vista frontal alrededor de 0 º y 180 º y una vista lateral para el resto de los ángulos. Salvo algunas excepciones (hay parámetros que no son visibles en la vista lateral, como el punto que une las piernas), en ambos casos, los parámetros del modelo se obtienen de la misma manera.

Observemos la figura 3.5. A modo de ejemplo del tipo de análisis que podemos

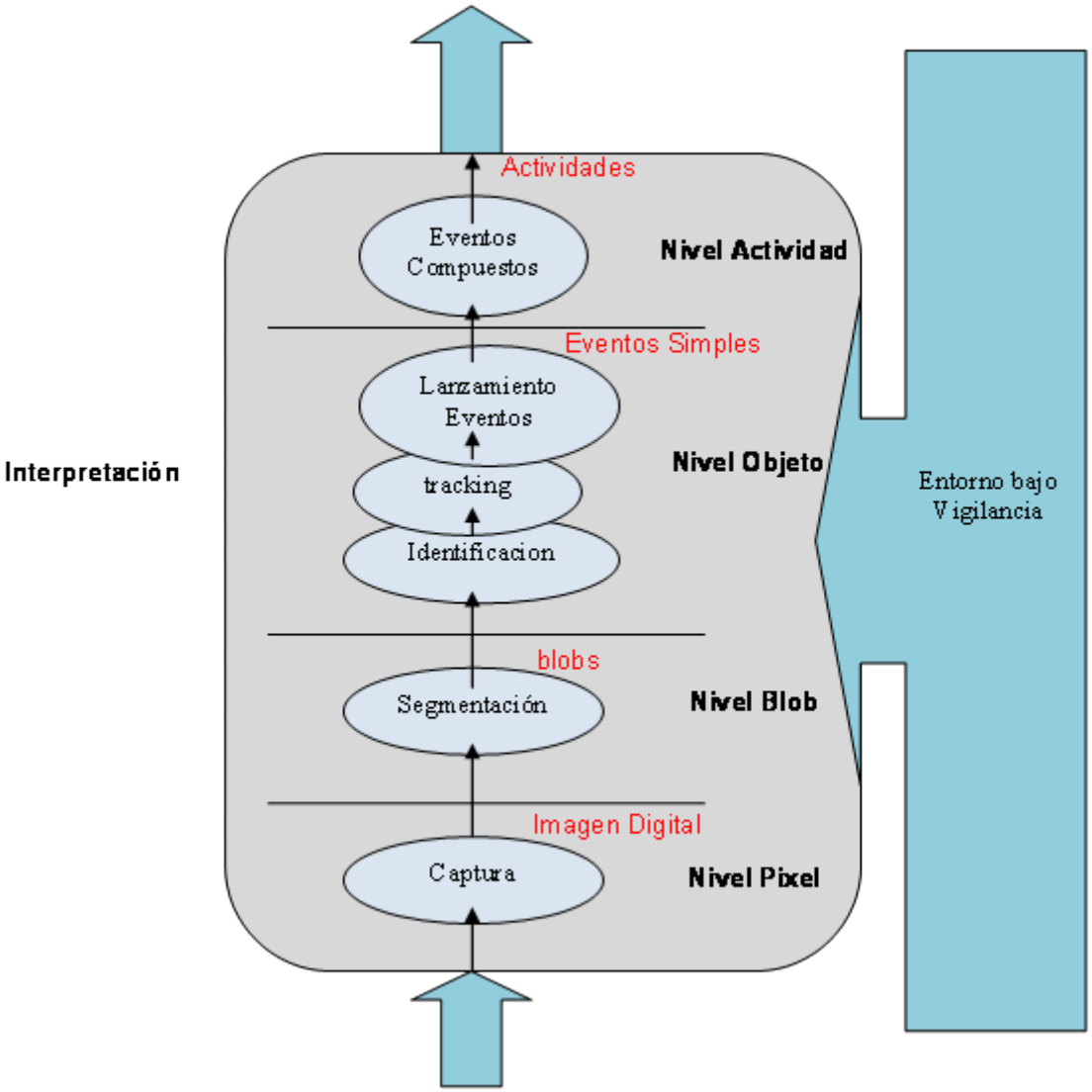


Figura 3.3: Estructura bottom-up detallada de los diferentes niveles de descripción de la escena.

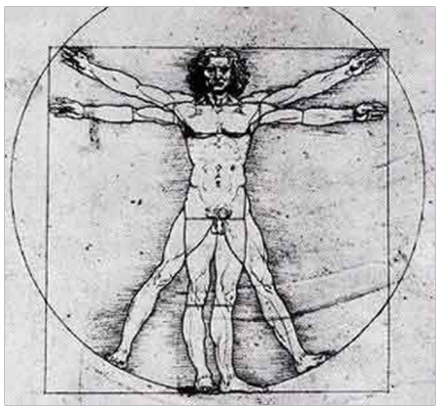


Figura 3.4: Hombre de Vitruvio

realizar con esta definición del modelo de bloques BB6-HM, se aprecia que los bloques B_1 y B_2 presentan una anchura muy superior para el caso frontal que para el lateral, o que los cambios en el tamaño de los bloques B_3 y B_4, con el movimiento de los brazos, serán mayores en el caso lateral que en el frontal. Por el contrario, los bloques B_5 y B_6 son más anchos para el caso lateral, si bien esto no va a ocurrir en todo momento, ya que la amplitud irá aumentando o disminuyendo de acuerdo con el movimiento de las piernas, mientras que para el caso frontal prácticamente apenas variarán. Otro ejemplo representativo podría consistir en la localización de las manos, en este caso nos fijamos solamente en los bloques B_3 y B_4, las manos se localizarán en los puntos intermedios más extremos horizontalmente de estos dos bloques, pudiendo estar en el bloque B_3 cuando el humano esté en movimiento o en B_4 tanto cuando está en reposo como cuando está en movimiento.

Esta división por bloques presenta cuatro propiedades muy importantes:

- Nos permite centrar nuestra atención en las regiones de interés e ignorar el resto.
- De la altura y anchura de los bloques, de las relaciones entre bloques y de su evolución temporal se puede extraer mucha información de utilidad.
- El modelo es invariante respecto del tamaño del humano: independientemente de la altura de éste, sea niño o adulto, o de la distancia con respecto a la cámara, los bloques se corresponden muy aproximadamente con las mismas partes del cuerpo humano.
- Esta descomposición en bloques es genérica y se puede aplicar a otro tipo de objetos, no sólo a humanos, configurando el número adecuado de bloques.

Cualquier información que pueda proporcionar la descomposición en bloques del modelo puede ser importante para el análisis de las escenas. Interesa, por tanto, describir características espacio-temporales del objeto, tales como el tamaño, la forma, el desplazamiento, etc. Como se comentó anteriormente, el modelo BB6-HM divide al humano en un grupo de seis bloques. Estos bloque van a venir caracterizados por un conjunto de parámetros básicos o primarios, que posteriormente se podrán combinar para definir nuevos parámetros secundarios orientados a la descripción de eventos concretos. En la sección 3.2.1 se describirán los parámetros primarios y puntos significativos de la silueta que van a caracterizar los bloques y que pueden verse en la figura 3.6. Y en la sección 3.2.2 describiremos los parámetros secundarios usados en este trabajo y otros que se pueden obtener para analizar el comportamiento de los bloques.

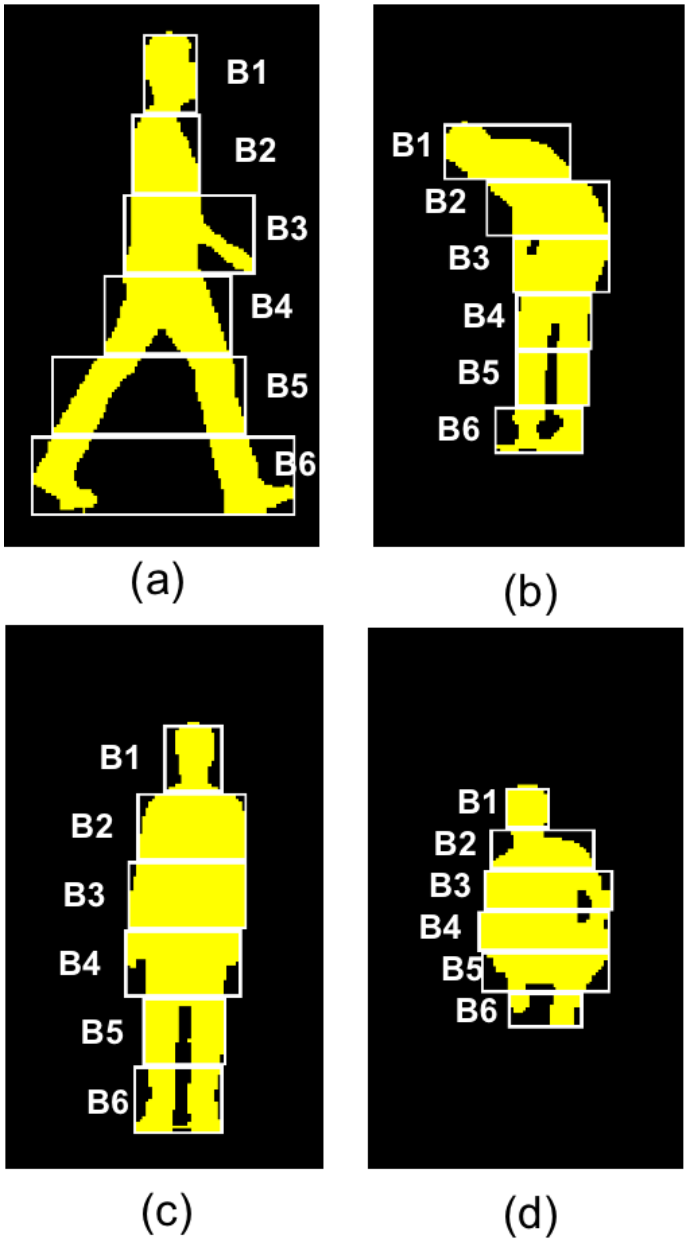


Figura 3.5: Modelo de bloques en vista frontal y lateral.

3.2.1. Parámetros primarios y puntos significativos

Para obtener los bloques es necesario, en primer lugar, obtener los puntos superior e inferior, P_C y P_{inf} respectivamente, del blob definido por la región de la imagen que se asocia al objeto de interés, en este caso, al humano en cuestión. Estos puntos proporcionan la altura del conjunto de bloques, H_T, que permite dividirlo en los diferentes bloques, B_i, $i = 1, .., 6$. Todos los bloques tienen la misma altura, $H_{B_i} = H_T/6$.

A continuación se obtienen los puntos del extremo izquierdo y derecho de la silueta localizada en cada bloque B_i. Con estos puntos extremos se define un nuevo parámetro: la anchura de cada bloque, W_{B_i}. La anchura máxima del conjunto de bloques se define como, W_T. Otros dos puntos significativos son el centro de masas del blob (CM) y el punto de unión de las piernas en la silueta, P_Λ, el cual sólo es significativo en el caso de vista lateral. A partir del centro de masas, podemos definir el eje de simetría, que es la línea vertical que pasa por el centro de masas. El eje de simetría divide los bloques horizontalmente en dos partes, lo que podemos caracterizar por la anchura del bloque a cada lado dicho eje, los parámetros W_{L_i} y W_{R_i}.

Calculando la intersección del blob con cada bloque, se calculan los puntos asociados a las manos, P_{H1} y P_{H2}, y a los pies, P_{F1} y P_{F2}. Los puntos asociados a las manos se sitúan en la intersección de la silueta con los lados verticales de los bloques B_3 o B_4, el que tenga la intersección más alejada del eje de simetría. Los puntos asociados a los pies se definen como los de las manos pero en el bloque B_6. Por último, la posición de las manos y los pies se describirá por medio de dos ángulos, θ y α :

- El ángulo θ formado por el punto superior del blob P_C y los puntos asociados a los pies P_{F1} y P_{F2}.

- El ángulo α formado por la línea que pasa por los puntos asociados a las manos, P_{H1} y P_{H2}, y la horizontal.

3.2.2. Parámetros secundarios

Partiendo de estos puntos significativos y parámetros primarios, que describen características espaciales básicas del modelo de bloques, se define un conjunto de parámetros secundarios relacionados con variaciones temporales o con relaciones de proporción entre ellos. Para estudiar cambios temporales es necesario disponer de

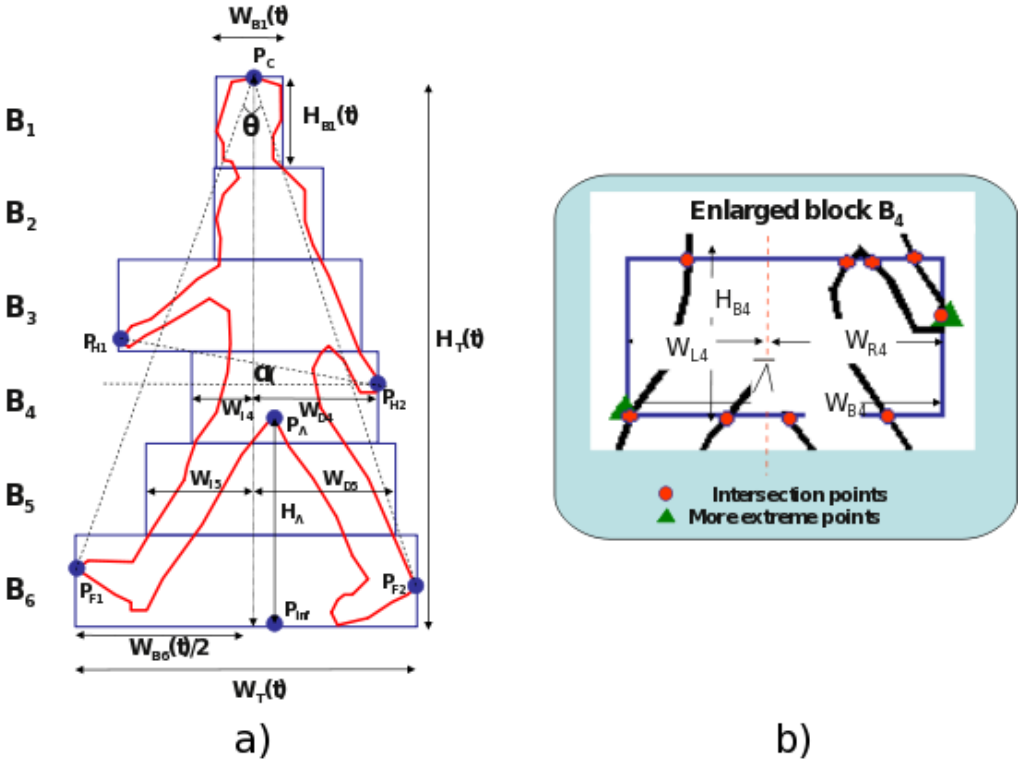


Figura 3.6: En la imagen de la izquierda se representa el modelo BB6-HM con los puntos más significativos y parámetros primarios: (a) vista general; (b) vista aumentada del bloque B_4. En la figura (a), además de los seis bloques ya definidos, se representa la altura del conjunto de bloques H_T, la anchura máxima del conjunto de bloques W_T, el punto de unión de las piernas en la silueta P_Λ y los puntos superior e inferiores (P_C y P_{inf}) del humano. En la figura (b), se hace una ampliación del bloque B_4 con el fin de apreciar los puntos de corte con el bloque y los parámetros W_{L_i} y W_{R_i}. En la imagen de la derecha se representa el modelo BB6-HM con un mayor número de parámetros.

una secuencia de imágenes, las secuencias se componen de fotogramas o *frames*[1] tomados a intervalos regulares de tiempo. El frame se indicará en las fórmulas que siguen a continuación mediante el símbolo t, la variable temporal. Para simplificar la nomenclatura se elimiará de las expresiones su dependencia temporal siempre que sea posible, esto es, salvo cuando esté implicado algún frame precedente $t - T$.

Estos son los parámetros secundarios que forman parte del modelo BB6-HM:

- Change_in_width, CW^T_{Wi} , (vector de cambio en la anchura):

$$CW_i^T = \frac{W_{Bi}(t)}{W_{Bi}(t-T)},\ \ i = 1, .., 6$$

donde cada componente contiene la relación entre el ancho del bloque en el frame t y el T-ésimo precedente $t - T$ para cada bloque B_i.

- Change_in_mass_centre, ΔCM^T,(vector de cambio en el centro de masas). Se define el cambio en sus dos componentes en la imagen, x e y:

$$\Delta CM_X^T = (CM_x(t) - CM_x(t-T))$$

$$\Delta CM_Y^T = (CM_y(t) - CM_y(t-T))$$

donde t representa el frame en el instante actual, $t - T$ representa el T-ésimo frame precedente y se pueden usar diferentes valores de T en función del problema. CM_x *y* CM_y y son las coordenadas x e y del centro de masas respectivamente.

- Directional symmetry vector, DS_i, (vector de simetría direccional):

$$DS_i = \frac{W_{Li}}{W_{Ri}},\ i = 1, .., 6$$

donde cada componente representa las proporción entre las anchuras de las partes del bloque B_i a la derecha y la izquierda del eje de simetría.

- Symmetry, S_i, (vector de simetría):

$$S_i = \frac{min(W_{Li}, W_{Ri})}{max(W_{Li}, W_{Ri})},\ i = 1, .., 6$$

donde cada componente representa la relación de simetría entre las anchuras de las partes del bloque B_i, a la derecha y a la izquierda del eje de simetría.

- Height_crutch, HC, (relación de altura total a altura del punto que une las piernas):

[1] Usaremos para la palabra fotograma su traducción al inglés frame, por ser usual en la literatura del área.

$$HC = \frac{H_T}{H_T - H_\Lambda}$$

donde H_Λ, se calcula como la distancia vertical desde el punto P_Λ al P_{inf}.

- Swinging feet coefficient, S_f, (coeficiente de balanceo de pies):

$$S_f = \frac{max(W_{L5}, W_{L6}) + max(W_{R5}, W_{R6})}{H_T}$$

este parámetro se utiliza para detectar actividades relacionadas con la postura. Si se asume que los pies son los puntos extremos de los bloques B_6 o B_5, este coeficiente es una medida de la apertura de las piernas.

- Swinging-hands coefficient, Sh, (coeficiente de balanceo de manos):

$$S_h = \frac{max(W_{L3}, W_{L4}) + max(W_{R3}, W_{R4})}{H_T}$$

este es un parámetro muy importante cuando se estudia la periodicidad, el tipo de movimiento y diferentes acciones asociadas con el movimiento de brazos.

Es muy importante destacar que esta lista es ampliable. A partir de los parámetros primarios y de los secundarios que se van definiendo, se pueden definir nuevos parámetros orientados a caracterizar distintas situaciones. Por ejemplo, si lo que se quiere es analizar la relación entre los bloques B_4 y B_5, se puede crear un nuevo parámetro que considere la relación que existe entre ambas anchuras y denominarlo, por ejemplo, W_{B4_B5}, el cual estaría definido por el cociente entre la anchura del bloque B_4 y B_5, es decir, $W_{B4\_B5} = W_{B4}/W_{B5}$. Este parámetro podría ser útil, por ejemplo, para diferenciar si se llevan los brazos pegados al cuerpo, si se lleva algún objeto, etc.

3.3. Información proporcionada por el modelo

El modelo de humanos BB6-HM permite caracterizar el movimiento de un humano en una secuencia de vídeo y detectar situaciones de interés para la tarea de vigilancia. Cada una de las situaciones reconocidas se describe como una función de los parámetros del modelo (parámetros primarios y secundarios).

Habrá casos en los que se analizarán simplemente las propiedades espaciales de la silueta en un instante t, mientras que en otros es necesario realizar un análisis espaciao-temporal. Los modelos para la descripción de las situaciones detectadas pueden ser estáticos o dinámicos.

La figura 3.7 describe el esquema general de reconocimiento de eventos. En una primera etapa, se obtiene una descripción del humano ("Human Description") que contiene parámetros constantes y variables. Los parámetros constantes constituyen

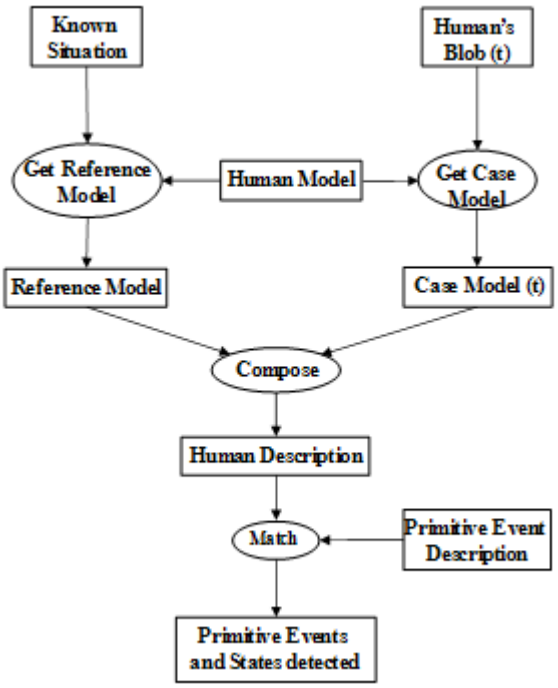


Figura 3.7: Diagrama del sistema para reconocer eventos primitivos de secuencias de vídeo con el modelo humano. Los *eventos primitivos* vienen predefinidos en el conjunto del sistema y se determinan a partir de cambios de estado de atributos visuales los cuales pueden utilizarse para la descripción de actividades y eventos más complejos.

el *modelo de referencia,* pues definen las características del ser humano en una situación conocida y actúan como una referencia para caracterizar la situación temporal del ser humano. Por ejemplo, la altura del humano de pie, H, es una constante que permite saber si el humano está erguido o sentado.. Los parámetros variables forman el *modelo del caso* y caracterizan al ser humano en un instante específico t. Estos parámetros se obtendrán de las características de los bloques. Los eventos de interés ("Primitive Event Descriptions") se caracterizan por las relaciones espacio-temporales de un subconjunto de los parámetros del modelo. Por lo tanto, la detección de los eventos consiste de una simple comparación de patrones ("match"), lo que permite definir un sistema modular y fácilmente extensible.

La información que podemos extraer del análisis con el modelo de bloques se puede clasificar en los siguientes tipos y será analizada en las siguientes secciones:

- Información sobre la localización de las partes del cuerpo.
- Información sobre el movimiento del humano: velocidad, dirección respecto a la cámara y periodicidad del movimiento.

- Reconocimiento de situaciones primitiva*s: postura y especiales, como oclusiones y acarreo de objetos.*
- Reconocimiento de la persona por la manera de andar ("Gait").

Esta información proporcionada por el modelo es la que se utilizará en el nivel de actividades para reconocer actividades de interés para la tarea de vigilancia. En la figura 3.8 se muestra un ejemplo del cronograma de eventos y estados generados en diferentes instantes de tiempo para una secuencia en la que se deja un objeto. En la parte superior de la figura, separada por la línea roja, se muestran los eventos que tienen lugar en la escena y que ocurren en un instante puntual. Así, en un instante de tiempo determinado t_1, una persona aparece en la escena con un objeto en la mano, con lo cual en el instante t_1 se tendrán dos eventos primitivos, por un lado el evento "entrar" y por otro el evento humano llevando objeto". Ambos se producen mientras la persona esta andando, con lo cual, en el cronograma (parte inferior de la fig. 3.8 debajo de la línea roja), aparecerá el estado "andando" como activado. Después se mantiene durante todo el intervalo de tiempo (t_1, t_2) andando hasta que decide en el instante t_2 parase, con lo que al producirse el evento "parar" el estado "andando" se desactiva. Durante todo este tiempo se ha estado llevando el objeto, por lo que el estado "llevando obj" se ha mantenido activo. En el instante t_3 se produce el evento de "agacharse" mientras sigue en el estado de "llevando obj", puesto que aunque no anda, el objeto aún no lo ha dejado. Después de este instante se ha producido la dejada del objeto en el instante t_4 que ha dado lugar al evento "no objeto detectado", a partir de este instante se pasa al estado "no llevando obj". Después de dejar el objeto, la persona se ha levantado, con lo cual ha dado lugar al evento "levantarse" en el instante t_5 mientras sigue en el estado "no llevando obj". Después en el instante t_6 decide empezar a moverse con lo cual se produce el evento "comenzar a andar" lo que hace que de nuevo se active el estado "andando". Durante el siguiente intervalo de tiempo (t_7, t_8) continua andando sin llevar ningún objeto para finalmente salir de la escena en el instante t_8.

3.4. Localización de las partes del cuerpo

Uno de los objetivos del modelo es identificar la posición de las diferentes partes del cuerpo, pues esto permite realizar un seguimiento específico de las partes del cuerpo y reconocer ciertas actividades. Para cada parte del cuerpo se analizarán

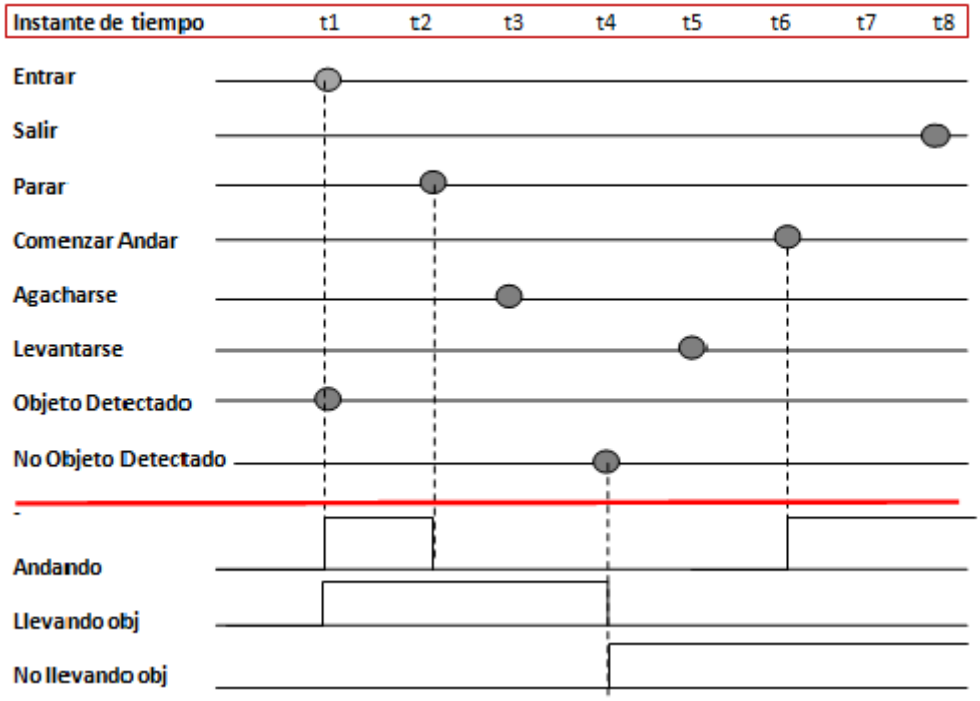


Figura 3.8: Ejemplo de eventos generados a partir de la secuencia segmentada de dejar un objeto. Se muestra un ejemplo de un cronograma de eventos generados, en diferentes instantes de tiempo, a partir de la secuencia segmentada para la situación de dejar un objeto. En la parte superior de la figura, separada por la línea roja, se muestran diferentes eventos que tienen lugar en la escena y que ocurren en un instante puntual y en la parte inferior diferentes estados producidos por los diferentes eventos.

unos bloques determinados. En concreto, en reposo (humano de pie, erguido y con los brazos hacia abajo) conocemos la posición las diferentes partes del cuerpo. En movimiento la situación se complica, por lo que será necesario analizar los bloques adyacentes a aquellos donde se encuentran las partes de interés en situación de reposo.

Así, las manos se corresponden de manera usual (no levantando los brazos, etc) con el bloque B_3 en situación de reposo cuando no hay movimiento de los brazos y con el B_4 cuando si lo hay. De igual manera, los pies se corresponden con los bloques B_5 en el caso de que se esté realizando movimiento y con el B_6 en caso contrario. La situación de la cabeza es diferente ya que indistintamente haya movimiento o no, siempre ocupa en situaciones normales (por ejemplo, no haciendo el pino) el bloque B_1 o B_2 si el individuo está con los brazos levantados.

Si queremos ser más precisos e identificar exactamente dónde están las manos, los pies o la cabeza, podremos utilizar las siguientes definiciones. Los puntos correspondientes a las manos P_{H1} y P_{H2} se definen como los puntos más extremos del bloque B_3 o B_4 dependiendo de la situación en que se encuentra, es decir, si está en reposo, lo usual es que estos dos parámetros se sitúen en el bloque B_4, mientras que si no lo está, estos oscilen entre el bloque B_3 y B_4. El movimiento es similar al de un péndulo que va subiendo y bajando según se avanza en el movimiento, lo que hace que la pertenencia a los bloques varíe. Para los pies se definen son los puntos P_{F1} y P_{F2}, los cuales se definen como los puntos extremos de los bloques B_5 o B_6 dependiendo de si está en situación de reposo o en movimiento. En posición de reposo y estando de pie, el bloque a considerar será el B_6 mientras que en movimiento los bloques considerados variarán entre el B_5 y B_6, es una situación similar a lo que ocurre en el caso de las manos. Otro punto a considerar para la localización de las partes del cuerpo es el P_{HD}, el cual se define como el punto extremo superior del bloque B_1 y que se corresponde con la cabeza cuando no se tienen los brazos levantados o el punto medio del punto de intersección superior del bloque B_2 con el bloque B_1 cuando un brazo o los dos están por encima de la cabeza. En la fig. 3.9, se muestran los puntos correspondientes a la cabeza, manos, pies y P_Λ. En la Tabla 3.1 se detallan las reglas heurísticas descritas, las cuales permiten asociar las partes del cuerpo con los bloques correspondientes de acuerdo a la situación estática o de movimiento. Estas definición son válidas tanto para el caso frontal como el lateral.

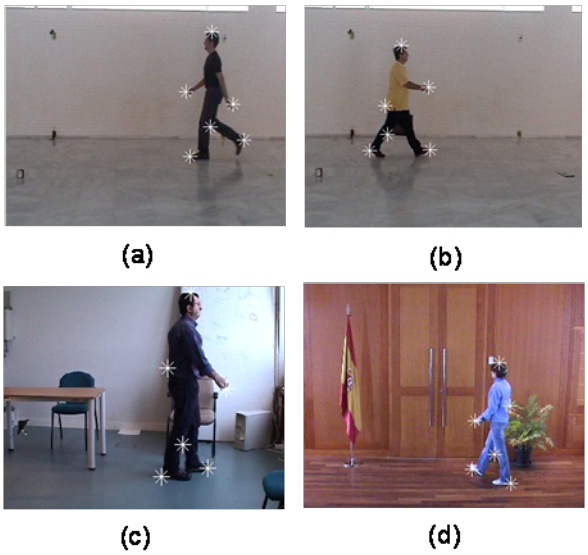


Figura 3.9: Secuencia usada para la localización de las partes del cuerpo. Los puntos PH_D, PH_1, PH_2, PF_1, PF_2 y P_Λ están marcados con *.

PART	DEFINITION
HANDS (P_{H1}, P_{H2})	Extreme right and left points of blocks B_3 or B_4: in the position of repose in B_4 and when there is movement in B_3 or B_4.
FEET (P_{F1}, P_{F2})	Extreme right and left points of blocks B_5 or B_6: in the position of repose in B_6 and when there is movement in B_5 or B_6.
HEAD (P_{HD})	Upper extreme point of block B_1 without the arms raised or midpoint of upper intersection points of block B_2 when arm or arms are raised over the head.
TORSO /BACK	Block situated immediately below the block to which the head belongs.

Tabla 3.1: Posición habitual de algunas las principales partes del cuerpo (manos, pies y cabeza) en el modelo de bloques (bloques con los que se corresponden), dependiendo de la situación en la que se encuentra el humano: en movimiento (andando, corriendo,... etc) o en situación estática.

3.5. Información sobre el movimiento del humano

En esta sección analizaremos tres características relacionadas con el movimiento: la velocidad, la dirección y el ángulo respecto a la cámara y la periodicidad del movimiento.

3.5.1. Análisis de la velocidad y la dirección del individuo

Para determinar la velocidad del desplazamiento del humano vamos a proceder a clasificar el movimiento en tres categorías: normal, lento y rápido. Para su identificación a partir de los parámetros del modelo, utilizaremos una combinación de definiciones basadas en el desplazamiento del blob en la imagen y la variación del ángulo θ.

Analisis a partir del desplazamiento del blob en la imagen.

La determinación de si un movimiento está clasificado como lento o rápido está basada en la comparación de dos cotas que se ajustarán de manera experimental:DIST_RAPIDO (Cota del movimiento normal) y DIST_LENTO (Cota del movimiento lento). Definamos "D" como la distancia euclídea que existe entre el punto medio del bloque donde se encuentra la cabeza, $(X_m,Y_m)\epsilon\ B_1$ entre un frame y el anterior. De modo que el movimiento es lento si "D" menor que la cota DIST_LENTO, el movimiento es normal si D está entre DIST_LENTO y DIST_RAPIDO y el movimiento es rápido si D es superior a DIST_RAPIDO. Cuando la distancia "D" es aproximadamente igual a cero el movimiento es nulo y por tanto esta parado.

Sea $(X_m,Y_m)\epsilon\ B_1$el punto medio del bloque donde se encuentra la cabeza. Se define la distancia D como:

$$D=\sqrt{(X_m-X_{m-1})^2+(Y_m-Y_{m-1})^2}$$

$$\text{si}\begin{cases} D < DIST\_LENTO & \rightarrow Lento \\ DIST\_LENTO \leq D \leq DIST\_RAPIDO & \rightarrow Normal \\ D \geq DIST\_RAPIDO & \rightarrow Rápido \end{cases}$$

$$\text{si } D\simeq 0 \rightarrow Parado$$

Análisis a partir de la variación del ángulo θ

Además de la obtención de estas cotas, se ha utilizado un modelo similar a un compás (fig. 3.11) para determinar el tipo de desplazamiento. Teniendo en cuenta el compás, se determina que el ángulo θ que forma la silueta de un humano entre la cabeza y los pies varía con el movimiento. Se puede determinar experimentalmente, mediante el estudio de este ángulo, que existe variación en el movimiento de uno lento a uno más rápido. Una apertura mayor del ángulo, en general, implicará una velocidad en el movimiento también mayor(, lo que se estudiará con el bloque B_6). Por tanto, este ángulo θ, da información acerca del tipo de movimiento:

Además de esto, el ángulo θ da también información del tipo de desplazamiento. Para ello se definen los puntos (X_i,Y_i) , (X_d,Y_d) ϵ B_6 como los puntos inferiores extremos del bloque B_6 pertenecientes a los pies y el punto medio del bloque donde se encuentra la cabeza, $(X_m,Y_m)\epsilon$ B_1 entre un frame y el anterior. También se define el punto de corte con el segmento determinado por las coordenadas (X_m, Y_m) y (X_i, Y_i), denominado como (X_c, Y_c). Con todos estos puntos se forma un triángulo (fig. 3.10) que permite determinar el ángulo θ. Para realizar la clasificación del movimiento se definen dos nuevas cotas asociadas a movimiento rápido, θ_f, y lento, θ_s.

Sean los puntos (X_i,Y_i) , (X_d,Y_d) ϵ B_6; los puntos inferiores extremos del bloque B_6 pertenecientes a los pies.

$D = \sqrt{((X_d - X_i)^2 + (Y_d - Y_i)^2)}$;
$A = \sqrt{((X_d - X_m)^2 + (Y_d - Y_m)^2)}$;
$C = \sqrt{((X_m - X_i)^2 + (Y_m - Y_i)^2)}$;
$B = (C^2–D^2 + A^2)/2 * C$;
$H = \sqrt{(A^2–B^2)}$;

$$Tg(\theta) = H/B;$$

De las pruebas realizadas en esta tesis y para los casos analizados se determinó heurísticamente que en el siguiente rango de valores se cumplían las siguientes condiciones:

$$\theta = \arctan(\theta);\ \text{si}\ \begin{cases} \theta < \theta_s & \rightarrow Lento \\ \theta > \theta_f & \rightarrow Rápido \\ \theta \geq \theta_s\ \text{y}\ \theta \leq \theta_f & \rightarrow Normal \end{cases}$$

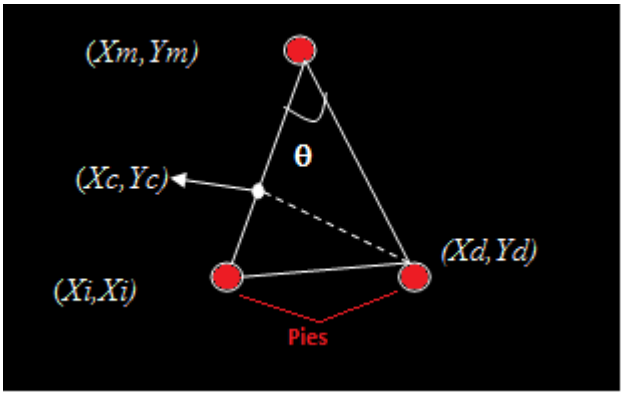


Figura 3.10: Coordenadas para la determinación del ángulo θ

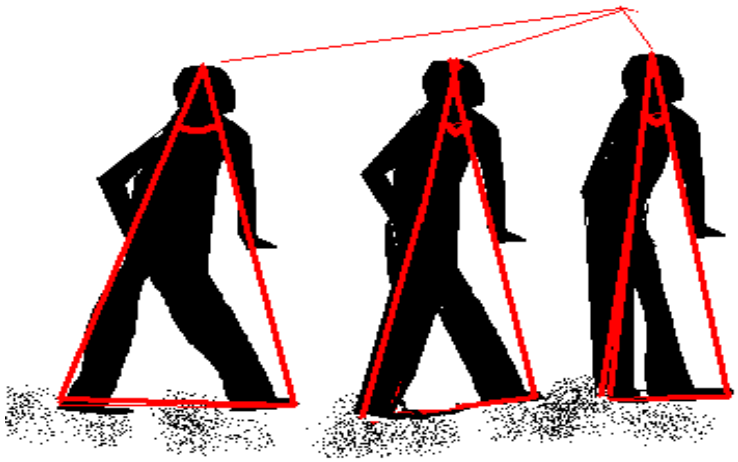


Figura 3.11: Ángulo entre cabeza y pies, θ, según el desplazamiento.Las líneas rojas definen la apertura del ángulo de modo que la figura de más a la izquierda tiene un ángulo mayor y por tanto mayor apertura de piernas, mientras que la de más a la derecha define un ángulo menor y por tanto menor apertura en las piernas. Esto proporciona información sobre el tipo de movimiento en el desplazamiento.

Análisis de la dirección del movimiento

En cuanto a la dirección el avance, Av, define la posición respecto a la cámara, cuyo valor se distingue entre: derecha, izquierda, acercándose y alejándose.

Sea $(X_{m-1}(1),Y_{m-1}(1))\epsilon\ B_1$ el punto medio del bloque donde se encuentra la cabeza en el frame anterior y $(X_m(1),Y_m(1))\epsilon\ B_1$ en el actual. Entonces:

$$\text{si}\begin{cases} X_m(1) > & X_{m-1}(1) \rightarrow Av = derecha \\ Y_m(1) < & Y_{m-1}(1) \rightarrow Av = alejamiento \\ X_m(1) < & X_{m-1}(1) \rightarrow Av = izquierda \\ Y_m(1) > & Y_{m-1}(1) \rightarrow Av = acercamiento \end{cases}$$

3.5.2. Análisis de la periodicidad

Un movimiento se dice periódico cuando se repite a intervalos regulares de tiempo. Hay que destacar que, para poder analizar la periodicidad del movimiento, es necesario definir previamente la escala a la que se desea realizar el análisis. El modelo BB6-HM podría utilizarse para comprobar el comportamiento periódico de algunos de sus parámetros y ser objeto de estudio. Esta sección contiene únicamente resultados del análisis realizado con el objetivo de mostrar las posibilidades del modelo, sin embargo esta periodicidad no se ha evaluado de forma cuantitativa y no se utiliza posteriormente para determinar ningún tipo de parámetro adicional, quedando su uso para trabajos futuros.

Dicho lo anterior, para ejemplificar este comportamiento periódico se ha tomado el parámetro HC desde una vista lateral del individuo, mostrando resultados cada cuatro frames. De acuerdo con esto, se han obtenido un conjunto de gráficas que permiten visualizar de una manera sencilla la periodicidad o aperiodicidad del movimiento en distintas situaciones. En la fig. 3.12 se representa la evolución del parámetro HC cuando el humano anda de forma continua y siguiendo un movimiento periódico y uniforme en el segmento de secuencia dado. En la parte superior de la figura, aparece una representación continua del movimiento que describe el humano y en la parte inferior una aproximación discreta simplemente con el fin de visualizarlo más claramente.

Se puede apreciar que la frecuencia de la señal periódica será la frecuencia del paso del individuo, lo que también nos da una información que permite caracterizar el paso y, por tanto, a la persona asociada el mismo. Así en la figura 3.13 se representa un ciclo del paso de dos humanos distintos desde que comienza con la piernas juntas

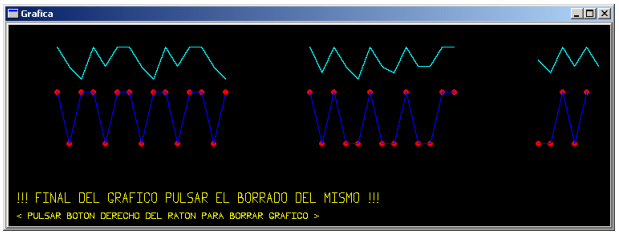


Figura 3.12: Periodicidad del movimiento. En la parte superior de la figura, aparece una representación continua del movimiento mediante el parámetro HC que describe el humano. En la parte inferior, se hace una aproximación discreta que se ha realizado tomando los cuatro puntos (en rosa en la imagen) más extremos y uniéndolos, simplemente con el fin de que se vea más claro

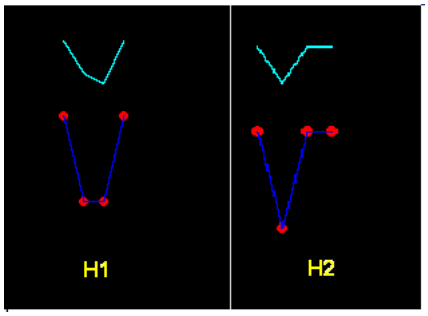


Figura 3.13: Ciclo del paso de dos humanos diferentes (H1 y H2). En la imagen de la izquierda se representa el paso de un humano (H1) en la secuencia desde que comienza con la piernas juntas y cerradas, hasta que termina el paso de la misma manera. La parte inferior representa la aproximación discreta. De igual modo, se representa en la izquierda la misma información pero para el humano (H2).

y cerradas, hasta que termina el paso de la misma manera.

En la fig.3.14 se muestra la evolución del ángulo θ (ángulo formado entre los pies y la cabeza) a lo largo de los 20 frames de dos secuencias distintas. La fig3.14a) muestra el ejemplo de una persona con movimiento periódico: anda moviendo sus brazos y piernas de manera uniforme, repetida y continuamente. Mientras en fig.3.14.b) muestra un movimiento que no sigue un patrón regular y por lo tanto es no periódico. Si se observa detenidamente la gráfica de la fig.3.14.a), se aprecia que cada cierto número de frames (aprox 7-8), el valor del ángulo θ alcanza 30^{o}, para luego descender de manera gradual hasta alcanzar un valor aproximado oscilante, entre 15 y 18^{o}. Es decir, el ángulo va variando de manera periódica a lo largo del transcurso de la secuencia. Sin embargo, en fig.3.14.b, los valores no siguen un patrón periódico. Según dicha figura, varían en un rango entre 15^{o}-40^{o}, sin ningún tipo de cadencia.

Este análisis de periodicidad del movimiento, permitiría en determinadas situaciones, saber si la persona está realizando paradas intermitentes o siguiendo un

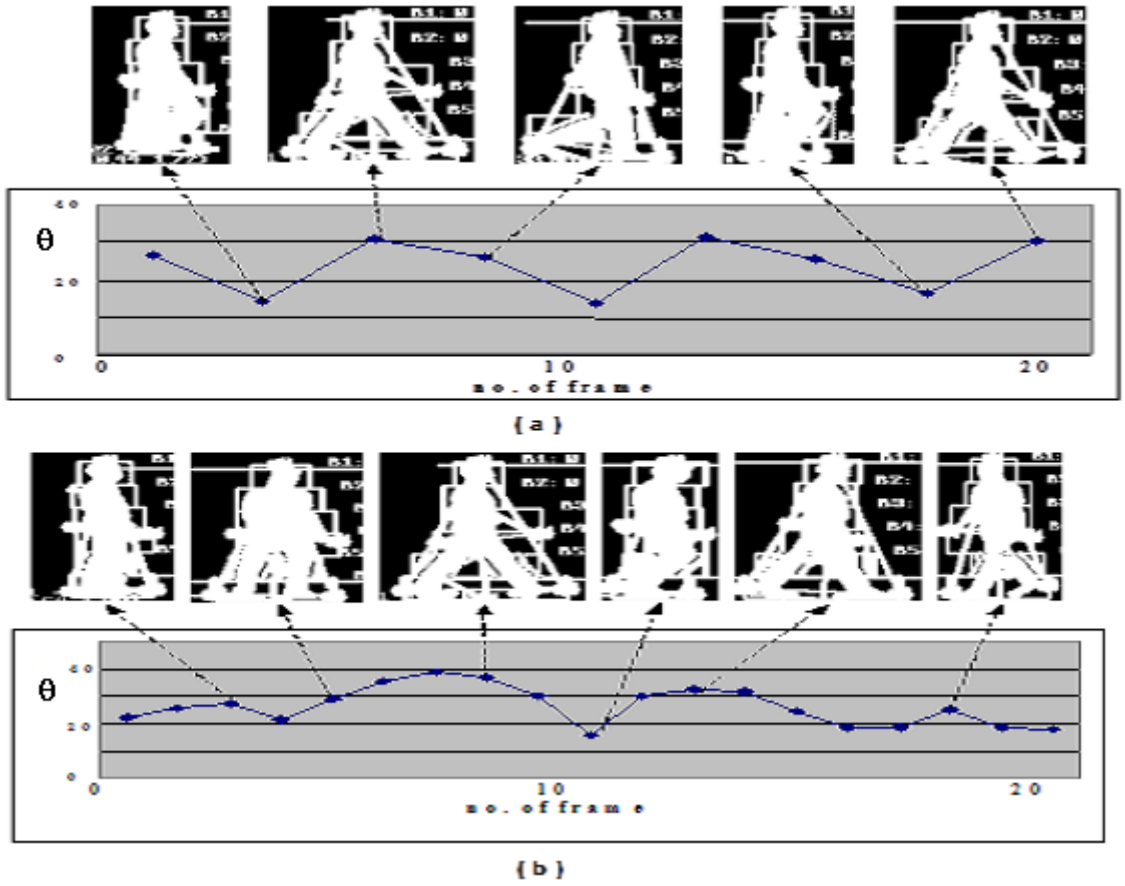


Figura 3.14: Gráfica de la periodicidad del movimiento: (a) Movimiento Periódico; (b) Movimiento No-Periódico.

movimiento regular, lo cual en temas de vigilancia puede ser bastante relevante para la detección de situaciones anómalas.

3.5.3. Análisis del ángulo del individuo respecto a la cámara

Aunque los parámetros del modelo de humanos BB6-HM son independientes de la vista, en muchas ocasiones los parámetros sólo son significativos en en caso de vista lateral o vista frontal. Esto hace que sea muy interesante conocer el ángulo que forma el individuo respecto a la cámara para conocer la confianza con la que podemos utilizar los valores obtenidos de los parámetros o incluso seleccionar el mejor modo de reconocer un cierto evento.

Dada la importancia de este ángulo dentro del sistema, el análisis del ángulo que forma el individuo respecto a la cámara se realizará mediante un sistema clasificador ajustado mediante aprendizaje supervisado. Para el ajuste del clasificador se utilizarán distintos algoritmos de aprendizaje supervisado con ejemplos tomados de la

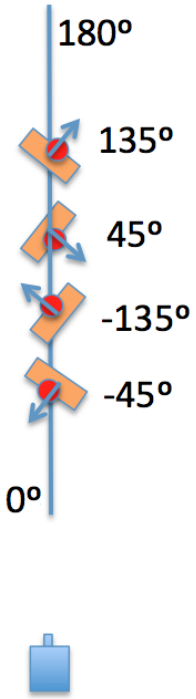


Figura 3.15: Ángulo de una persona respecto a la cámara.

base de datos de CASIA (ver sección 4.1.1.2).

La figura 3.15 muestra las limitaciones a la hora de reconocer este ángulo. Dado que el modelo sólo utiliza información de la silueta y con una sola cámara no tenemos información de profundidad, no se puede distinguir si una persona está de frente o de espaldas, e incluso resulta complicado distinguir si mira hacia la izquierda o hacia la derecha, pues la representación de la silueta en el plano de la imagen es similar. El reconocimiento de si es vista frontal o vista lateral resulta mucho más sencillo.

3.6. Reconocimiento de eventos y situaciones primitivas

En esta sección se describen los eventos y situaciones primitivas reconocidas a partir de los parámetros del modelo BB6-HM. Cada uno de los eventos y situaciones detectadas se basa en un subconjunto de los parámetros del modelo de bloques seleccionado en función del análisis específico de la situación. En todos los apartados de esta sección el subindice "i" corresponde al número de bloque.

Figura 3.16: Secuencia de frames segmentadas con el humano realizando la acción de agacharse. Al realizar la acción de agacharse se aprecia en las figuras que el conjunto de bloques se ha comprimido dando lugar a un altura total menor que la que tenía inicialmente. La anchura total también aumenta, así como, los bloques desde elB_1 al B_5 para este caso.

3.6.1. Agacharse

La acción de agacharse implica la compresión de los seis bloques y la altura total del blob disminuye:

$H_T(t) < H_T(t-T)$, donde H_T es la altura total del blob y se han considerando tres frames anteriores (T=3) lo que significa 0,32 sg. Teniendo en cuenta la evolución de la altura, se comprueba que las alturas se van reduciendo a medida que se el humano se agacha y permanece constante, con una altura total menor que H, mientras el humano permanece agachado. Pudiera darse que una persona se fuese agachando a medida que avanza en la escena, pero en este análisis se ha considerado que la persona apenas avanza o está parada del todo.

En caso de vista lateral, además de lo dicho anteriormente, generalmente se cumple que la anchura total de los bloques en las frames anteriores es superior

a las posteriores. En la fig. 3.16se aprecia una secuencia en la que una persona aparece agachándose. Al realizar la acción de agacharse se aprecia que el conjunto de bloques se ha comprimido, dando lugar a una altura total menor que la que tenía inicialmente. La anchura total también aumenta, así como la de los bloques desde elB_1 al B_5 para este caso. La cabeza puede dejar de ser el punto más alto del conjunto de bloques, pasando a ser la espalda.

Algunas condiciones para estar agachándose son:

- $W_{B_{i-1}}(t-1) > W_{B_i}(t)$, $i = 1.,5$
- $H_T(t-3)$ y $H_T(t-2)$ y $H_T(t-1) > H_T(t)$
- Está parado (la posición (x,y,z) respecto a la cámara no ha variado) o movimiento lento (3.5)

  Además, si la vista es lateral.
- $W_T(t-3)$ y $W_T(t-2)$ y $W_T(t-1) > W_T(t)$
- $DS_i \leqq 1$ Indica que se agacha hacia la derecha y en caso contrario a la izquierda. Con $i = 5$.

3.6.2. Levantarse

Esta situación es la opuesta a agacharse. Aquí también se involucran los seis bloques, como en el caso anterior. Los seis bloques pasan a tener una altitud : $H_T(t)$ $> H_T(t-T)$, donde H_T es la altura total del blob y se han considerando tres frames anteriores (T=3) lo que significa 0,32 sg. Puesto que la altura total aumenta, todas y cada una de las alturas de los bloques H_i aumentan también.. Se almacena también la altura de cada uno de los bloques de los anteriores frames, así como la altura total y se comprueba que la altura de cada bloque va aumentando a medida que se avanza en la escena. Al igual que en el caso anterior, pudiera darse que una persona se fuese levantando a medida que avanza en la escena, pero en este análisis heurístico se ha considerado que la persona está parada mientras se levanta o avanza lentamente mientras efectúa esta acción.

También se comprueba si la vista es lateral o frontal. En caso de vista lateral además de lo dicho anteriormente se cumple que para el caso lateral la anchura total de los bloques en las frames anteriores es inferior a las posteriores.

Por tanto algunas condiciones son:

- $W_{B_{i-1}}(t-1) < W_{B_i}(t)$, $i = 1.,5$

- $H_T(t-3)$ y $H_T(t-2)$ y $H_T(t-1) < H_T(t)$
- Está parado (la posición (x,y,z) respecto a la cámara no ha variado) o movimiento lento (3.5)

  además si la vista es lateral.
- $W_T(t-3)$ y $W_T(t-2)$ y $W_T(t-1) < W_T(t)$

3.6.3. Saltar (verticalmente)

Saltar (fig. 3.17) es otro estado en el se consideran los seis bloques. Como ya se comentó, la acción de saltar es casi un estado combinado de la acción de agacharse y levantarse, sólo que se realizará a mayor velocidad. Por tanto, al tratarse de una combinación de ambas, los parámetros a estudiar son los mismos, sólo que realizados de manera consecutiva. Se produce una elevación del centro de masas (ΔCM_M^T), respecto del suelo que viene determinado por el punto P_{inf}. Cuando se produce una elevación sobre el suelo el parámetro P_{inf} sufre una variación en su coordenada y y apenas en la coordenada x respecto a las frames anteriores. Para el caso del salto se iban analizando cada dos frames consecutivas (la actual y la anterior). Esto podría confundirse con la acción de alejarse en la escena, sin embargo, en este caso la altura total de los bloques H_T no se reduce, si no que se mantiene igual e incluso aumenta un poco (al saltar se realiza un estiramiento del cuerpo). Lo mismo sucede con W_T, el cual se reduce también al alejarse en la escena. Las coordenadas del centro masas siguen el mismo patrón que el punto P_{inf}

Por tanto, se cumplen las mismas condiciones que en saltar y levantar añadiendo además las siguientes:

- $P_{infX}(t-1) \simeq P_{infX}(t)$
- $P_{infY}(t-1) < P_{infY}(t)$
- $H_T(t-1) \simeq H_T(t)$ y $W_T(t-1) \simeq W_T(t)$
- $CM_x(t-1) \simeq CM_x(t)$
- $CM_y(t-1) < CM_y(t)$

3.6.4. Brazos levantados

Aquí se quiere detectar si una persona ha levantado sus brazos. La altura total va aumentando de manera gradual entre frames consecutivas hasta llegar a incrementarse la altura total del humano en casi un $H_T/3$ (experimentalmente en las pruebas

Figura 3.17: Frame segmentada con el humano realizando la acción de saltar. En la imagen de la izquierda se comienza el proceso del salto al agacharse el humano con el fin de impulsarse, con lo cual los bloques en conjunto se comprimen y por consiguiente la altura total. así como, la disposición de los bloques. En la imagen de la derecha el humano ya ha efectuado el salto y se ha añadido una raya amarilla para distinguir que está por encima del suelo. En esta imagen los bloques se vuelven a estirar y la altura total aumenta. Hay que fijarse que en el bloque B_6los pies aparecen en posición inclinada formando una diagonal lo que hace que el tamaño dicho bloque varie.

realizadas) sobre la altura del humano. De modo que haciendo un seguimiento de los bloques B_1 y B_2se llega a la detección de esta situación.

Habría que distinguir la detección de este evento desde las dos perspectivas lateral y frontal, ya que los parámetros en ambos casos es diferente. La forma de elevar los brazos puede variar, dependiendo de si se elevan hacia delante del cuerpo o por los lados, pasando por los brazos extendidos, hasta llegar a tenerlos completamente estirados por encima de la cabeza.

En el caso frontal y si los brazos se elevasen por delante del cuerpo apenas se apreciaría la elevación hasta llegar a la altura del bloque B_2 . En el caso lateral y elevación de brazos por los lados tampoco se apreciarían variaciones para esta acción sólo cuando se llega a superar el bloque B_1. En ambos casos, a partir del bloque B_1, la altura total H_T empieza a aumentar gradualmente hasta que al final de la situación la altura total del humano en casi un 1/5 (experimentalmente en las pruebas realizadas) de la altura total.

En esta tesis se ha analizado el caso frontal y la elevación de los brazos se ha producido por los lados (fig. 3.18), con lo cual se han podido observar variaciones de más parámetros que las alturas. En este caso se observa que hay un significativo incremento en el ancho y alto de los bloques B_1 y/o B_2 en el actual frame comparado con los precedentes frames, y una amplia variación en el ángulo γ formado entre las

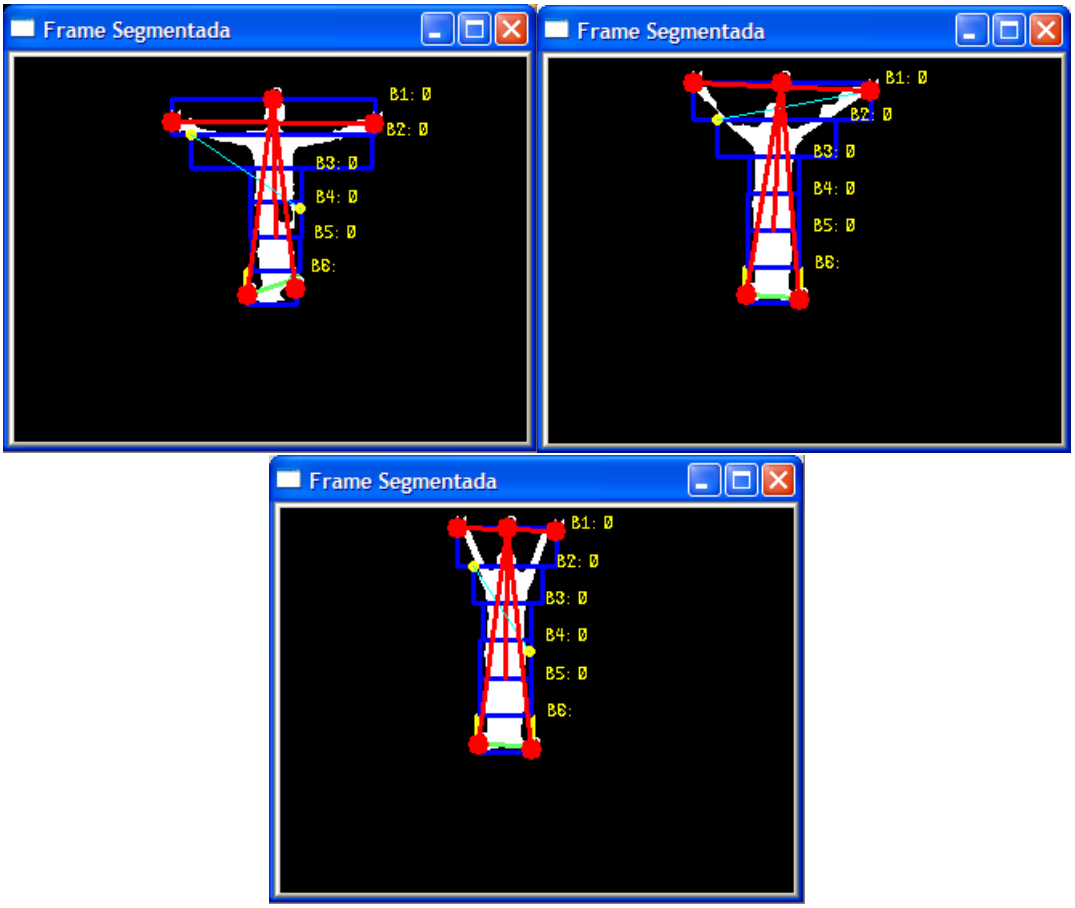


Figura 3.18: Secuencia de frames donde se realiza la acción de elevar los dos brazos desde la vista frontal y por al lateral del cuerpo. La secuencia, que comienza desde arriba a la izquierda hasta abajo, muestra el momento en que ya se ha pasado de los brazos extendidos y se sigue elevando hasta dejarlos estirados. En esta situación última el punto medio del bloque B_1 esta situado por encima de la cabeza.

manos con la cabeza, a medida que se va avanzando en la secuencia hasta llegar al momento en que los brazos pasan por la situación de brazos extendidos (sección 3.6.5), con lo cual el análisis es el mismo, pasando con un enorme aumento de la amplitud del bloque B_1 hasta elevarse por encima de la cabeza para después producirse gradualmente un decremento de la misma hasta el momento de pararse. Al superar el punto más alto del bloque B_1 la altitud del individuo empieza a estar por encima de la suya. Considerando tres frames consecutivas y el caso analizado en esta tesis:

- Paso por los brazos en extendidos o en "cruz".
- A partir de un instante de tiempo $t \rightarrow W_{B_1}(t) \geq W_T(t)$ y $\gamma(t) < 0$.
- $\forall Bi$, $i = 1.,6$; $B_{i_{final}} > B_{i_{inicial}}$
- $H_{T_{final}} \simeq 6 \times \frac{H_{T_{inicial}}}{5}$

3.6.5. Brazos extendidos

En esta situación se trata de un caso particular del anterior, ya que sigue una elevación de los brazos pero en vez de subirlos por encima de la cabeza permanecen como máximo a la altura de los hombros (fig3.19). El análisis de esta situación tiene sentido desde el punto de vista frontal, ya que desde el lateral no se podría apreciar. Es el caso de una persona que se pone con los brazos en "cruz", hasta llegar a incrementarse la anchura total del humano a aproximadamente la altura del humano, H. La anchura total va aumentando de manera gradual entre frames consecutivas, de modo que va aumentando desde el bloque B_4 hasta el B_2, donde alcanza la máxima amplitud.

Por tanto, estos son los bloques implicados en la detección de esta situación. La variación de la anchuras para los bloques B_2, B_3 y B_4 en tres frames consecutivas sería:

- $W_{B_4}(t-2) > W_{B_3}(t-2) > W_{B_2}(t-2)$
- $W_{B_3}(t-1) > W_{B_4}(t-1)$ y $W_{B_3}(t-1) > W_{B_2}(t-1)$
- $W_{B_2}(t) > W_{B_3}(t)$ y $W_{B_2}(t-1) > W_{B_4}(t-1)$
- $W_T(t) \simeq 2 * W_T(t-2)$
- $\gamma(t-2) \ll \gamma(t)$ y se comprueba experimentalmente en las pruebas realizadas para los casos analizados que:$\gamma(t) \simeq 180^{\text{o}}$

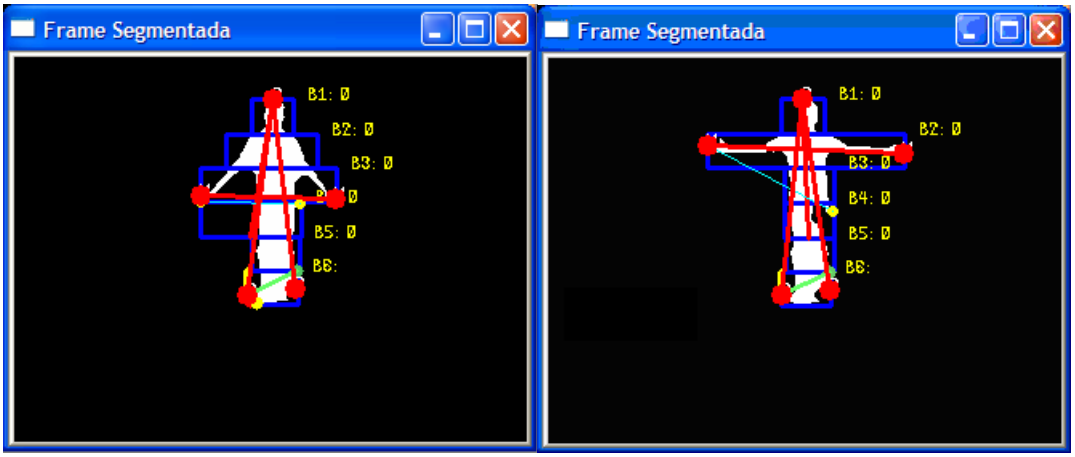


Figura 3.19: Frames segmentadas con el humano realizando la acción de extender los brazos. En la imagen de la izquierda el humano empieza a realizar la acción de elevar los brazos ampliándose los bloque B_3 y B_4. En la imagen de la derecha aparece una persona que ha extendido los brazos por completo y como consecuencia el bloque B_2aparece con una anchura muy superior a lo normal. Además, se observa, de igual modo, que el bloque B_1aparece con un tamaño más grande de lo normal ya que por el lado de la izquierda no esta ajustado a la cabeza, esto es porque ha pillado en la división de los bloques un trozo del hombro que le hace que aumente.

- Al final del proceso cuando se llega a la extensión de los brazos: $W_T = W_{B_2}$

3.6.6. Tumbarse

Los seis bloques pasan a tener una altitud $H_{Td} < H_{Ta}$, donde H_{Td} es la anchura total del conjunto de bloques después de agacharse y H_{Ta} antes de hacerlo. Puesto que la altura total se reduce, todas y cada una de las alturas de los bloques H_i se reducen también. A medida que se van analizando frames consecutivas se aprecia esta reducción, mediante el análisis de los diferentes bloques. La diferencia entre la acción de agacharse y tumbarse radica en el hecho de que los bloques se comprimen aún más. En la anchura total ocurre todo lo contrario, es decir, aumenta de gran manera respecto a la que tenia en instantes anteriores $W_{Td} > W_{Ta}$.

3.6.7. Giros

Se determinan los puntos más extremos de los pies, de modo que se pueda ver cual es la punta y cual es el talón. Para ello se analizan los puntos de corte con el bloque B_6, de modo que se trazan rectas verticales desde estos puntos de corte hasta encontrarse con el bloqueB_5 dependiendo de si encuentra puntos entre medias

Figura 3.20: Puntos de los pies según la orientación.

o no se determina la punta. En cualquiercaso, el bloque B_6 cambiará de forma y de tamaño concreto a medida que se va realizando el giro y la punta va variando. Los parámetros asociados a su detección, son la altura y la anchura del bloque B_6 y el ángulo θ. En determinadas frames, el cambio puede no ser detectable, pero el hecho de utilizar una ventana de frames hace posible determinarlo.

En la fig.3.20, se aprecian los puntos de los pies, tanto los puntos delanteros como los puntos traseros. Utilizando las rectas obtenidas por regresión como aproximación del contorno de los pies, se establecen cuáles son las puntas de los pies y de los talones. Como se aprecia en la figura, en color amarillo, las rectas obtenidas dan una aproximación de las puntas o talones, ya que en estos casos, estas rectas no cortan o aproximan a ningún punto interior al bloque B_6. Una vez que se sabe lo que es (punta o talón), y teniendo constancia de la dirección del movimiento (izquierda o derecha) que llevaba el humano, se puede determinar cuál es la punta del pie.

3.6.8. Recoger objeto

Este es un evento que involucra en su detección a diferentes bloques según la manera de recoger el objeto, aunque el más importante en la detección suele ser el B_4. Tiene dependencia con las frames anteriores y es una situación que involucra a bastantes parámetros al realizarse las acciones de agacharse, coger un objeto y finalmente levantarse. Por tanto, se puede decir que este es un evento compuesto de otros tres. La acción de agacharse y levantarse se detectará del modo que se ha explicado en las secciones 3.6.1 y 3.6.2. La acción de recoger un objeto es una acción compleja ya que se puede realizar de varias maneras. En esta tesis se han analizado dos casos típicos.

Caso uno: se puede realizar estando sentado y recogiendo el objeto de lado, extendiendo el brazo de modo que se produce un aumento y posterior reducción de la anchura de los bloques B_2, B_3 y B_4, como de indica en la figura 3.21 superior. Para este caso, el punto P_{inf} permanece prácticamente invariante durante la acción de recoger un objeto y el ángulo α va aumentando en la extensión y reduciéndose en la contracción. En la figura se aprecia a un humano recogiendo un maletín, para

lo cual extiende el brazo y luego lo va contrayendo. El bloque B_4 aparece marcado en un círculo y relleno en color rojo para destacar el cambio de amplitud, aunque también los bloques B_2 y B_3 sufren una contracción.

Caso 2: se considera que el humano está de pie (fig.3.21 inferior) y se agacha con el fin de recoger algún objeto. En este caso todos los bloques se ven involucrados.

En ambos casos se ha analizado la frame actual con la anterior en un intervalo de 0,16 sg Se han considerado los casos tanto para la vista lateral como frontal. El análisis del caso estando sentado que se hace aquí es desde el punto de vista frontal. Y el caso de estar de pie que se contempla aquí es el de vista lateral.

Estado sentado:

- Agacharse (acción de sentarse)
- Extensión del brazo:
  - $W_{B_{i-1}}(t-1) < W_{B_i}(t)$, $i = 2.,4$ con $P_{inf}(t-1) \simeq P_{inf}(t)$
  - $\alpha(t-1) < \alpha(t)$
  - Se comprueba experimentalmente en las pruebas realizadas para los casos analizados que: $S_i \ll 0.5$, $i = 2.,4$, siendo i el número de bloque.
- Contracción del brazo:
  - $W_{B_{i-1}}(t-1) > W_{B_i}(t)$, $i = 2.,4$ con $P_{inf}(t-1) \simeq P_{inf}(t)$
  - $\alpha(t-1) > \alpha(t)$
  - Se comprueba experimentalmente en las pruebas realizadas para los casos analizados que: $S_i \simeq 1$, $i = 2.,4$
- Levantarse con el objeto en la mano.

  Estado de pie:
  - Agacharse
  - $W_{B_{i-1}}(t-1) < W_{B_i}(t)$, $i = 1.,6$ con $P_{inf}(t-1) \simeq P_{inf}(t)$
  - $\alpha(t-1) < \alpha(t)$
  - Se comprueba experimentalmente en las pruebas realizadas para los casos analizados que: $S_i \ll 0.5$, $i = 4.,6$
  - Levantarse con el objeto en la mano.

Una vez realizada esta acción (levantarse) se continuará con la acción de llevar objeto, la cual se explica con más detalle en la sección 3.6.10.

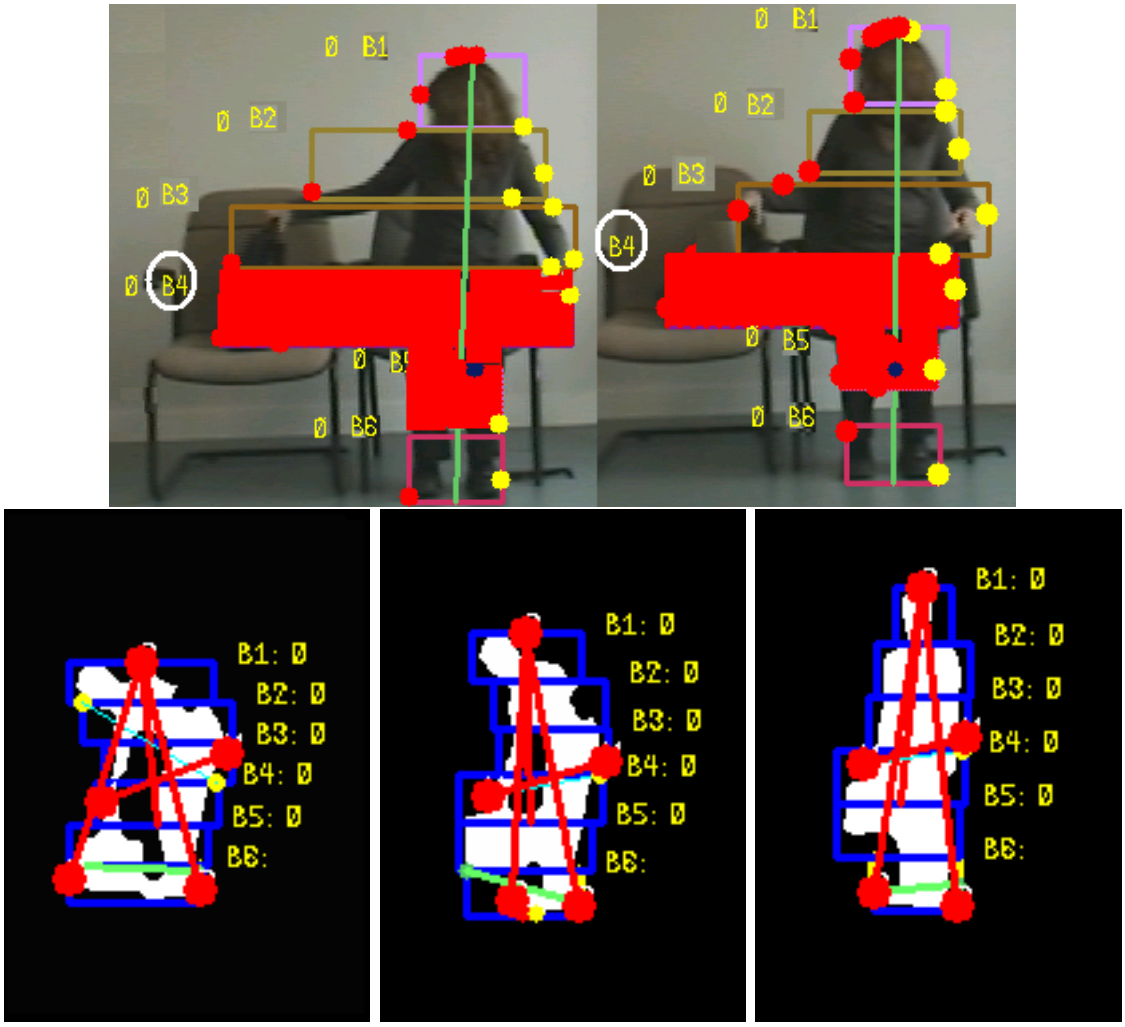


Figura 3.21: Arriba el humano esta sentando por lo que se recoge el objeto de lado extendiendo el brazo de modo que se produce un aumento y posterior reducción de la anchura de los bloques B_2, B_3 y B_4. Abajo el humano está de pie y se agacha con el fin de recoger algún objeto. En este caso todos los bloques se ven involucrados.

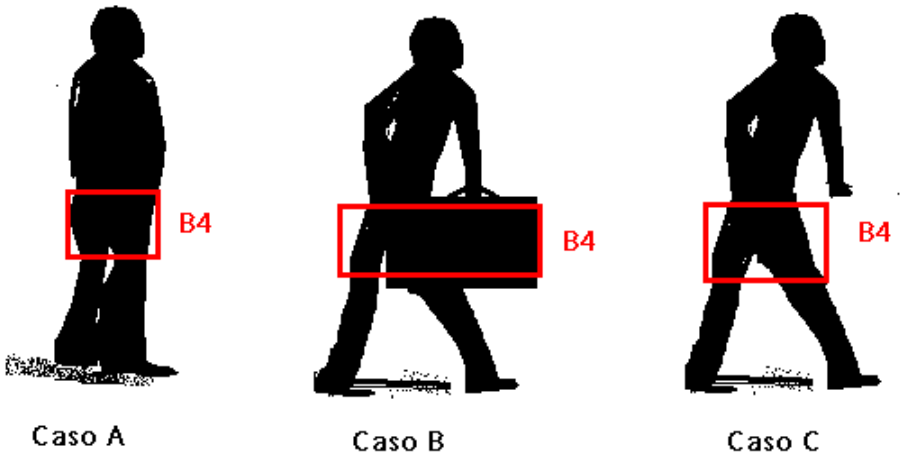


Figura 3.22: Posibles situaciones a la hora de llevar un objeto: (A) se podría pensar que cuando una persona lleva un objeto, normalmente, los brazos apenas se mueven, no hay balanceo.(B) Lleva un objeto y (C) No lleva nada.

3.6.9. Dejar objeto

Este evento es similar al caso anterior (recoger objeto) sólo que invirtiendo el proceso. Así, es un evento que involucra en su detección a diferentes bloques según la manera de recoger el objeto aunque el más importante suele ser el B_4. Tiene dependencia con las frames anteriores y es una situación que involucra a bastantes parámetros, al realizarse la acción de llevar un objeto, agacharse y finalmente levantarse. Por tanto se puede decir que este es un evento compuesto de otros tres. La explicación sería igual a la que se hizo en 3.6.8 sólo que invirtiendo el proceso.

3.6.10. Llevar objeto en la mano

En esta situación, se parte del hecho de que la persona lleva un objeto en la mano, por ejemplo un maletín. Para analizar esta situación, se parte de la figura 3.22, donde se representan 3 casos posibles. En el caso A), se podría pensar que cuando una persona lleva un objeto, normalmente, apenas hay balanceo en el movimiento de brazos , lo cual podría ser un factor a considerar a la hora de la detección. Sin embargo, hay que tener en cuenta que hay personas que apenas mueven los brazos al andar. Luego estarían los casos en los que sus manos parecen ocupadas y, por tanto, lleva un objeto (caso B) y el que sus manos están libres y por tanto no lleva nada (caso C).

Teniendo en cuenta que la posición en la cual un humano suele llevar el objeto es el bloque B_4, pues corresponde a la zona aproximada donde acaba el brazo, vamos

a analizar las diferencias en este bloque. En la figura 3.22 se aprecia claramente que existe una diferencia de anchura cuando lleva un objeto o cuando no. Luego este es un posible factor a considerar a la hora de evaluar la secuencia, pero no suficiente. Hay que pensar que, cuando una persona anda moviendo mucho los brazos, el bloque B_4 se amplia significativamente, sin embargo, el movimiento de los brazos es pendular, por lo que el bloque B_4 quedará reducido y será el bloque B_3 el que aparecerá aumentado. Irá así secuencialmente aumentando y disminuyendo en el caso de no llevar ningún objeto. Por tanto, se van a definir a continuación nuevos parámetros, que en base a numerosas pruebas, se obtienen valores numéricos que delimitan el valor de los mismos.

Condiciones de llevar objeto

Dado el frame actual, se considera la información de la altura y la anchura en el frame anterior en el instante $t-1$. Dichos valores serán informaciones relativas a cada bloque B_i.

De este modo se tiene almacenada la información:

$$H_i(t-1),\ W_i(t-1)\ \ con\ i=1,..,6.$$

Con estos valores se crean los coeficientes: AC_{43} y AC_{42}, que relacionan la anchura del bloque B4 con los bloques 3 y 2 respectivamente

de manera que:

$$AC_{43} = W_4/W_3; \quad AC_{42} = W_4/W_2$$

Se define el término: Abrupt Change AC (cambio_brusco), como el cambio que se produce cuando un bloque B_i, entre una "frame" y la siguiente, pasa de una amplitud a otra muy diferente. Por tanto el termino

$$AC = ((AC_{43}(t-1) - AC_{43}(t)\ ,\ (AC_{42}(t-1) - AC_{42}(t))$$

Con estos parámetros, se detectan situaciones en las que se deja o se coge un objeto. En las figuras siguientes se aprecia este hecho (fig. 3.23 y 3.24). En este caso se muestra una amplitud exagerada de los bloques B_2, B_3 y B_4, que sirve para determinar el valor de los coeficientes AC_1 y AC_2.

Lo anterior, permite identificar "cambios" en los bloques pero, además de ello, se necesita tener más parámetros que aseguren el hecho de que realmente el sujeto lleva un objeto. Por tanto, se definen dos parámetros más:

- Carriyng Objects, (γ), coeficiente de acarreo de objetos.

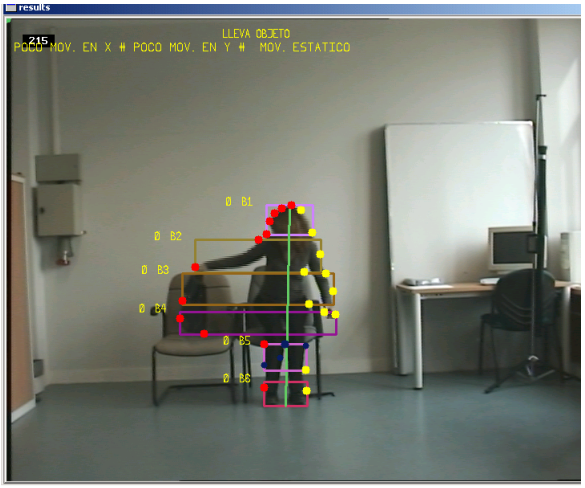


Figura 3.23: Situación de dejar un objeto. En este caso se muestra una amplitud exagerada de los bloques B_2, B_3 y B_4 que sirven para determinar el valor de los coeficientes AC_1 y AC_2.

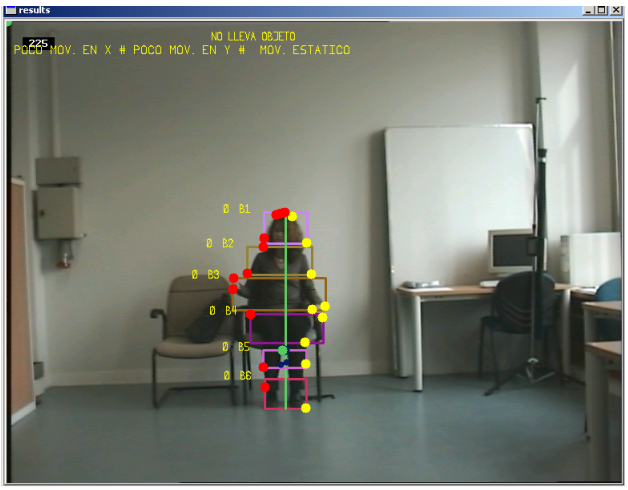


Figura 3.24: Situación de objeto dejado. En este caso el objeto ya ha sido dejado y por tanto en esta frame los bloques B_2, B_3 y B_4 han vuelto a su situación original y por tanto el valor de los coeficientes AC_1 y AC_2 han cambiado.

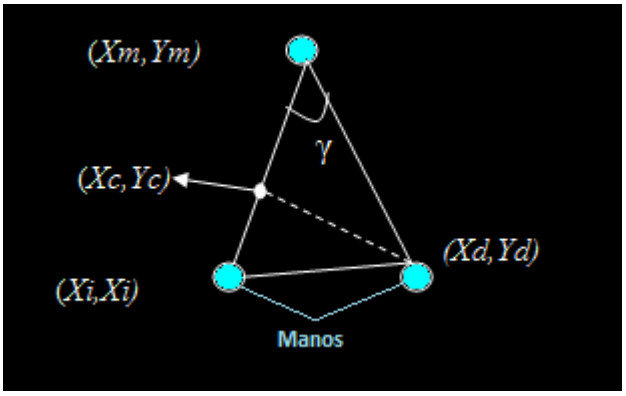


Figura 3.25: Coordenadas para la determinación del ángulo γ

- Height_Crutch (HC), relación de altura de la entrepierna.

Carriyng Objects (γ) es el ángulo que forma el punto medio del bloque donde se encuentra la cabeza (B_1 en situación normal) con los puntos más extremos de corte del blob con el bloque B_4 (normalmente estos puntos pertenecen a las manos). Se comprueba experimentalmente que cuando se lleva un objeto, la variación de este ángulo es muy pequeña a cuando si lleva. Cuando ocurre durante un conjunto de frames consecutivas se entiende que está llevando un objeto. Se considera, como mínimo, la información de los dos frames anteriores al actual.

Sea $(x_m, y_m) \in B_1$ el punto medio del bloque donde se encuentra la cabeza. Sean los puntos $(x_i, y_i); (x_d, y_d); \in B_4$, los puntos inferiores extremos del bloque B_4 y (x_c, y_c) el punto de corte con el segmento que une los puntos (x_m, y_m) y (x_i, y_i). Entonces, con estos tres puntos se forma el triángulo que define el ángulo γ (fig. 3.25) como:

$D = \sqrt{((Xd - Xi)^2 + (Yd - Yi)^2)};$
$A = \sqrt{((Xd - Xm)^2 + (Yd - Ym)^2)};$
$C = \sqrt{((Xm - Xi)^2 + (Ym - Yi)^2)};$
$B = (C^2 – D^2 + A^2)/2 * C;$
$H = \sqrt{(A^2 – B^2)};$

$$\gamma = arcTg(H/B);$$

En la figura 3.25se muestra un esquema de las coordenadas y de los puntos para entender mejor el significado de las mismas

En las figuras 3.26 y 3.27 se aprecia la diferencia en la variación del ángulo

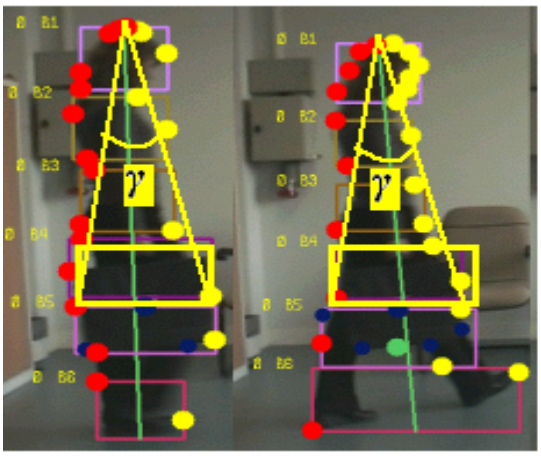


Figura 3.26: Ángulo con objeto. Variación del coeficiente Carriyng Objects en la situación en la que se lleva un objeto. En este caso la variación es muy pequeña y se ha obtenido, de manera experimental y con los casos estudiados en esta tesis, que el valor se sitúa entre $(0 \leq \gamma \leq 10^{o})$.

cuando se lleva o no se lleva objeto en las manos.

En el caso de vista lateral, otro parámetro que proporciona información sobre el acarreo de objetos es el parámetro HC. En la fig. 3.28 se observa la diferencia entre llevar un objeto o no en la altura del punto que une las piernas. En caso de vista frontal, se se produce una asimetría en el conjunto global de la figura que conforma los bloques, ya que aparece más gruesa que si no llevara objeto. En este caso, se produce un aumento de la anchura de los bloques que suelen estar cerca de las manos que es bastante característico. Normalmente serán los bloques B_4 y B_5 los que se vean afectados por esta situación y sirvan para determinar, considerándolos entre diferentes frames, si el sujeto lleva o no objeto. Esta situación se refleja en fig. 3.29.a, donde se observa el caso en que la persona no aparece en la escena de lado, sino frontalmente. En este último caso, el parámetroHC no es de mucha utilidad.

3.7. Detección de oclusiones

La oclusión de un humano se puede dar porque aún no haya entrado completamente en la escena o porque algún otro objeto estático o dinámico se interponga entre el humano y la cámara. Esto provoca una pérdida de información de eventos relacionados con el humano completo o con algunas de sus partes.

En el caso de oclusión por un objeto estático, que forma parte del escenario y que es conocido con anterioridad, es fácil detectar la oclusión, pues es posible realizar

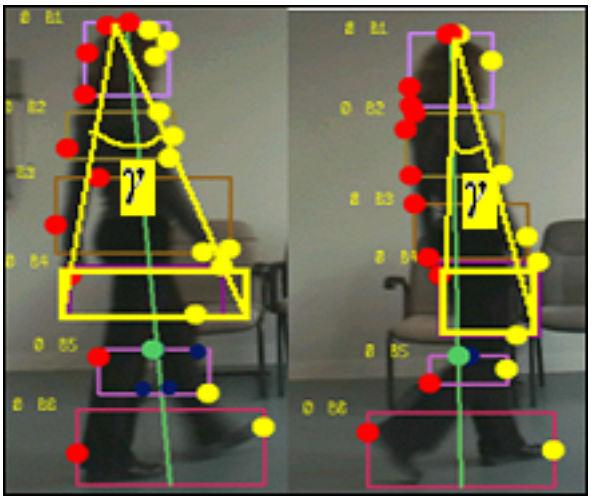


Figura 3.27: Ángulo sin objeto. Variación del coeficiente Carriyng Objects en la situación en la que no lleva un objeto. En este caso la variación es mayor que en el caso anterior y donde se ha obtenido, de manera experimental y con los casos estudiados en esta tesis, que el valor cumple ($\gamma \geq 10^{o}$).

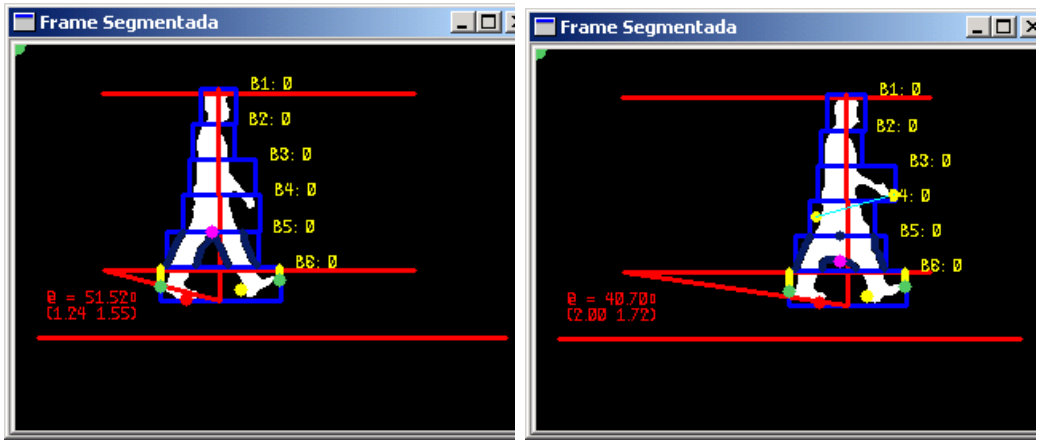


Figura 3.28: Coeficiente HC para el caso de no llevar objeto o llevarlo de lado. En la imagen de la izquierda se aprecia que la persona no lleva nada, tiene una altitud casi constante con respecto al suelo. El coeficiente HC se ha obtenido, experimentalmente y para esta tesis con las secuencias probadas, que varía entre 1 y 3,75. Mientras que en la imagen de la derecha donde lleva un maletín la distancia al suelo varia más y el coeficiente HC , tiene un valor diferente al caso anterior.

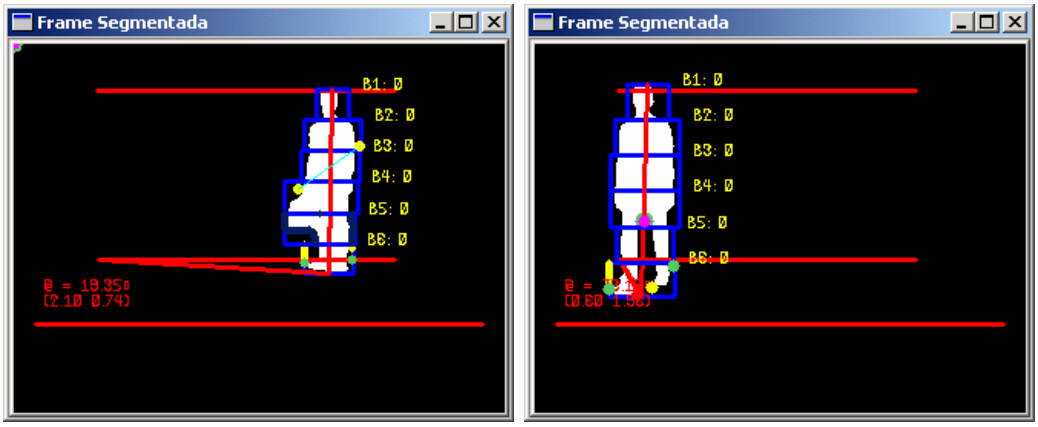


Figura 3.29: Asimetría y simetría del humano cuando lleva (a) y no objetos (b). En los casos frontales y para el caso de llevar un objeto en la mano, se observa en el análisis que se produce una asimetría en el conjunto global de la figura que conforma los bloques, ya que aparece más gruesa de lo que sería normal con respecto a la forma si no llevará objeto. En este caso, se produce un aumento de bloques característicos que suelen estar cerca de las manos. Normalmente serán los bloques B_4 y B_5.

un registro del escenario en el que se establezcan las coordenadas de los objetos que pueden ocluir a otros. Para analizar la situación de oclusión, se analiza el caso de una persona que avanza de modo lateral a través de la secuencia de imágenes donde se han marcado de manera manual los puntos extremos del objeto que ocluye. En este ejemplo, se supone que es un objeto rectangular con las esquinas representadas por los puntos $(xo, yo), (x1, y1), (x2, y2), (x3, y3)$. Inicialmente, la persona se aproxima a dicho obstáculo pero aún no hay oclusión (fig. 3.30). Más adelante, en las frames siguientes, el humano sigue avanzando hasta que ambos blobs iteractúan, es decir, la persona ha alcanzado el obstáculo. (fig. 3.31

Esta invasión por parte del humano de la zona de oclusión hace que el tamaño de los bloques que se tratan se reduzca hasta llegar, en determinados casos, a una eliminación total de los mismos. Esta situación marcará que hay una oclusión total en alguna parte del cuerpo.

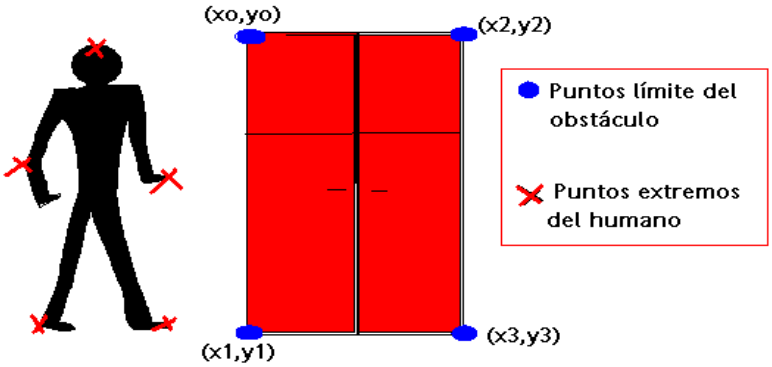


Figura 3.30: Humano fuera de la oclusión y al obstáculo con sus puntos marcados. Los puntos extremos del humano, representados con cruces, aún están separados de la posición del objeto. Los puntos azules representan los límites del obstáculo.

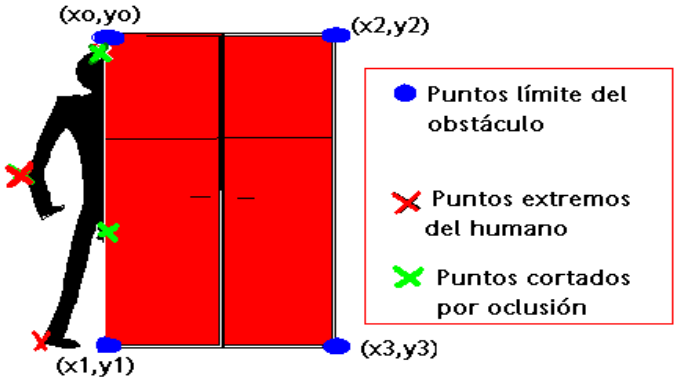


Figura 3.31: Humano en la oclusión. Se produce un corte de parte de su cuerpo. Ahora se tienen nuevos puntos, representados por cruces verdes, los puntos cortados por oclusión. Estos puntos marcan que se está invadiendo el espacio de la oclusión.

Sea la altura del bloque B_i:
$H_{B_i} = |x(min) - x(max)|$
y su anchura:
$W_{B_i} = |y(max) - y(min)|$

Sean $x(min)$, $x(max)$, los puntos del eje horizontal de B_i, mínimos y máximos respectivamente. Sea $y(max)$, $y(min)$ para el caso del eje vertical.

Sean los puntos (píxeles de la imagen) $x_j \in B_i$; con $x_j \in \mathbb{N}$; $j = 1, .., m.$; Y sea$Ext(i) = \{e_1,....,e_k\}$, $e_k \in \mathbb{N}$; tal que, $Ext(i) \subset B_i$ (conjunto de puntos extremos del bloque "i").

Sean los puntos extremos de la oclusión definidos como aquellos puntos más extremos del contorno que forma el obstáculo (puntos superiores izquierda y derecha, así como, puntos inferiores izquierda y derecha). (3.32):

Ocl = $\{(x_o, y_o),(x_1, y_1),(x_2, y_2),(x_3, y_3)\}$;

y Cocl un contorno cerrado de forma irregular (el hecho de representarlo con un contorno irregular se debe al hecho de que la forma no tiene porque ser una figura geométrica perfecta), que representa el blob del obstáculo delimitado por los puntos extremos, Ocl.
Entonces las condiciones de oclusión parcial de un bloque vienen dadas por:

1. ($H_{B_i} < H_{B_{(i-1)}}$ y $H_{B_{(i-1)}} < H_{B_{(i-2)}}$) ó ($H_{B_i} > H_{B_{(i-1)}}$ y $H_{B_{(i-1)}} > H_{B_{(i-2)}}$) algún $e_p \in Ext(i)$ y $e_p \in Ocl$; *con* $i = 1, .., 6.$

2. ($W_{Bi} < W_{B_{(i-1)}}$ y $W_{B_{(i-1)}} < W_{B_{(i-2)}}$) ó ($W_{Bi} > W_{B_{(i-1)}}$ y $W_{B_{(i-1)}} > W_{B_{(i-2)}}$); algún $e_p \in Ext(i)$ y $e_p \in Ocl$; *con* $i = 1, .., 6.$

A continuación se muestran algunos ejemplos de oclusiones totales o parciales de humanos. En la fig. 3.33, se aprecia el caso de un humano que ya está segmentado y que sufre una oclusión parcial con un obstáculo, el cual ocluye parte de los bloques B_3 y B_4. En la imagen se muestra como el obstáculo ha invadido parte del contorno del blob del humano. En esta situación, a medida que el humano ha ido adentrándose en el obstáculo, la anchura de los bloques B_3 y B_4 ha ido comprimiéndose cada vez más. Esto marca el hecho de que se ha producido una oclusión ya que de ninguna manera (a no ser que el problema sea que se ha hecho una mala segmentación), los bloques B_3 y B_4 pueden comprimirse.

En la figura (figs.3.34.a y 3.34.c) se aprecia el caso de un humano que ya está segmentado y se encuentra con un obstáculo, ello da lugar a una oclusión total de

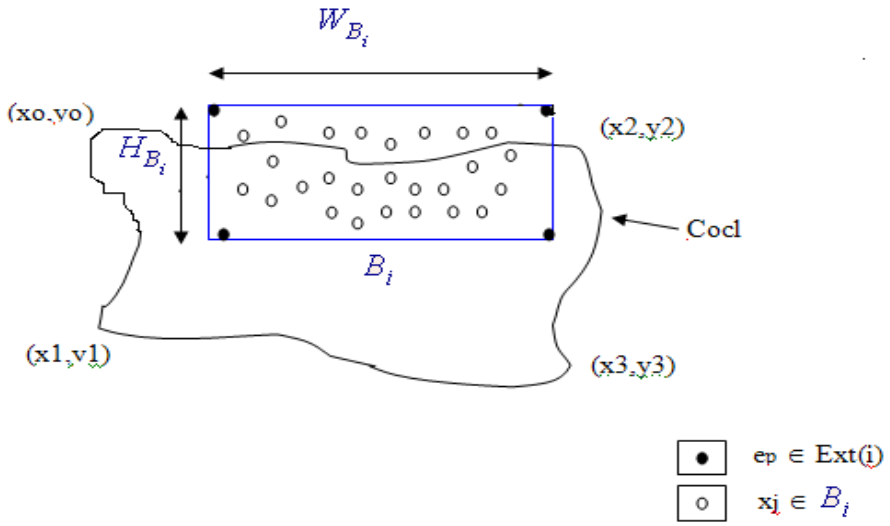


Figura 3.32: Representación de los puntos para la oclusión. En la imagen se representan los parámetros del bloque "i" involucrados en el proceso de oclusión. Los puntos $\{(x_o, y_o),(x_1, y_1),(x_2, y_2),(x_3, y_3)\}$, corresponden a los valores extremos del obstáculo. Cocl representa el contorno del obstáculo. Ext(i) son los puntos extremos del bloque "i".

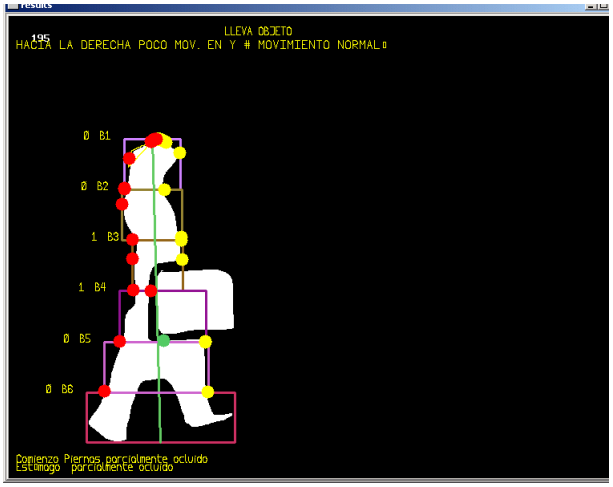


Figura 3.33: Oclusión parcial. Cuerpo obstaculizado a la mitad del mismo por una caja. En la imagen se muestra como el obstáculo ha invadido parte del contorno del blob del humano. En esta situación, a medida que el humano ha ido adentrándose en el obstáculo, la anchura de los bloques B_3 y B_4 ha ido comprimiéndose cada vez más.

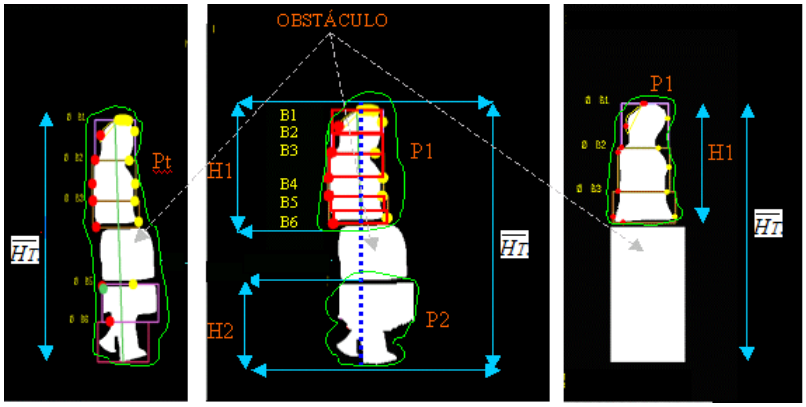


Figura 3.34: Casos del individuo segmentado. (a) Correcto (b) Incorrecto (c) Correcto. Se aprecia el caso de un humano que ya está segmentado y se encuentra con un obstáculo, ello da lugar a una oclusión total de bloques que ocluye totalmente el bloque B_4, en el caso (a), y, B_4, B_5, B_6 en el caso (c). Estas situaciones se consideran correctas. En el caso (b), el bloque B_4 desaparece, ya no debe de aparecer porque no proporciona ninguna información. Si no se hiciera se tomaría como que se están dando dos segmentaciones diferentes (quedaría partido en dos). El humano aparece con los seis bloques comprimidos, lo cual es una modelización incorrecta.

bloques que ocluye totalmente el bloque B_4 en el caso (a) y los bloques B_4, B_5, B_6 en el caso (c).

En el caso de oclusión total de un bloque, hay que volver a evaluar el modelo, eliminando el bloque que desaparece y tomando toda la información previa del modelo, antes de la oclusión, para ser capaz de mantener los parámetros anteriores antes de la entrada en la oclusión, tales como: altura, anchura, ángulos, proporciones, etc. Por tanto, para este caso se tiene que realizar un análisis del modelo anterior y plasmarlo de nuevo en otro modelo posterior que conserva las características de los anteriores pero que omite la información errónea asociada a la oclusión.

En el caso de la figura anterior, el bloque B_4 desaparece, ya no debe de aparecer porque no proporciona ninguna información. Si no se hiciera se tomaría como que se están dando dos segmentaciones diferentes (quedaría partido en dos). En la fig. 3.34.b, se ve este hecho, donde el contorno P_1 contiene los bloques contraídos y no la división correcta.

Se definen dos contornos para evaluar la oclusión parcial en caso de dividir al humano en dos fragmentos. Por un lado el contorno P_1, que es el contorno que contiene los puntos del blob segmentado del primer trozo en el que se ha partido el blog y P_2 que contiene los del segundo. Por tanto, sea $P_t = P_1, P_2$; como el conjunto formado por todos los puntos del blog y que se corresponde con la suma de los puntos de P_1 y P_2. Además se considera la altura, H_T, definida como la altura del

humano guardada en "frames" anteriores y la altura de cada uno de los conjuntos de puntos. Por un lado H_1 de P_1 y por otro lado H_2 de P_2 de modo que se cumple que la suma de ambas es menor a la total, es decir, $H_1 + H_2 < H_T$. Es importante tener en cuenta que los bloques deben distribuirse entre los dos contornos, de modo que $P_1 + P_2$ contendrán todos los puntos del blob y por tanto todos los bloques.

En el caso de que la oclusión parcial sólo divida al humano en un fragmento, sólo se consideraría el conjunto de puntos dado por P_1 y por tanto el conjunto total de puntos P_t sólo estaría comprendido por el P_1. En este caso la altura media H_T la altura del humano guardada en "frames" anteriores será mayor que la dada por el contorno P_1, es decir $H_1 < H_T$.

Por tanto se pueden tener dos situaciones diferentes:

1) Fig. 3.34.b

Sean los puntos $x_j \in P_1$; con $x_j \in \mathbb{N}$ y $j = 1, ..., m$. Y sean los puntos $x_k \in P_2$; con $x_k \in \mathbb{N}$ y $k = 1, ...l$; correspondientes a los dos contornos P_1 y P_2 segmentado. Se tiene pues el conjunto $P_t = P_1, P_2$; formado por todos los puntos de P_1 y P_2.

Sea la altura, $\bar{H_T}$, la altura del humano guardada en "frames" anteriores y sea H_1 la altura de uno de los conjuntos de puntos y H_2 el del otro tal que: $H_1 + H_2 < H_T$

2) Fig. 3.34.c

Sean los puntos $x_j \in P_1$; con $x_j \in \mathbb{N}$ y $j = 1, ..., m$. contorno P_1 segmentado. Se tiene entonces el conjunto $P_t = P_1$; formado por todos los puntos P_1. Sea la altura media H_T la altura del humano guardada en "frames" anteriores y sea H_1 la altura del conjunto de puntos tal que:

$$H_1 < H_T$$

3.8. Reconocimiento de personas por la forma de andar (Gait)

La detección, autentificación o reconocimiento de humanos siempre ha sido un tema constante de investigación. La revisión bibliográfica (subsección 2.3) demuestra el gran interés en obtener resultados computacionalmente fiables y en tiempo real

para construir sistemas el control de la identidad de las personas a distancia (en caso de lugares muy sensibles) o sin requerir la colaboración de la persona. Uno de las formas que se utiliza para conseguir esto es utilizar la capacidad de reconocer personas mediante la forma de caminar (gait). Reconocer una persona por la forma de caminar es una tarea muy compleja pues la forma de caminar se ve claramente afectada cuando se modifica la apariencia del individuo, sobre todo si se modifica la silueta por motivos tales como llevar puesto un abrigo o un sombrero, tirar de una maleta, etc.

Dado que realizar la tarea de reconocimiento del "Gait es compleja y difícil de describir de forma analítica, usaremos diferentes técnicas de clasificación basadas en aprendizaje supervisado para construir modelos estáticos y dinámicos que permitan reconocer a un individuo por la forma de caminar a partir de los parámetros del modelo BB6-HM. Se utilizará como entrada todos los parámetros del modelo y se seleccionarán los que proporcionan mayor capacidad discriminante a partir de métodos de aprendizaje. Para analizar la robustez del reconocimiento, se entrenará el sistema para reconocer personas con indumentarias distintas o que porten algún un maletín en las manos o una mochila a la espalda, ya que estos factores harán variar su apariencia.

Para la construcción de modelos estáticos para el reconocimiento del "Gait" se utilizarán diferentes clasificadores y se compararán los resultados obtenidos. Como subconjunto de entradas se obtuvo el vector

$$\left(\theta, \gamma, \Delta CM_M^5, H_4, H_5, W_4, W_2, W_3, W_5, W_6, S_2, S_4, S_6, S_f, S_h S_{B4\_B5}\right)$$

Por otro lado, se utilizarán modelos dinámicos mediante el uso de los modelos ocultos de Markov o HMM, de modo que se obtendrá un HMM por cada uno de los individuos de la escena y de ahí se seleccionará, bien el máximo (obteniendo un clasificador), bien una ordenación de los modelos más semejantes. Como parámetros de entrada se utilizarán, inicialmente, todos los parámetros del modelo para posteriormente ir ajustándolos hasta obtener un vector de parámetros más reducido o subconjunto del inicial que sea capaz de caracterizar al individuo. Al igual que en el caso de los estáticos se obtuvo un vector

$$\left(DS_6, DS_4, \theta, \gamma, \Delta CM_M^5, H_4, H_5, W_4, W_2, W_3, W_5, W_6, S_2, S_4, S_6, S_f, S_h S_{B4\_B5}\right)$$

Los dos modos de aprendizaje trabajan de forma diferente, ya que mientras en los métodos estáticos el clasificador obtiene como salida un único individuo seleccionado, en los métodos dinámicos se permite manejar múltiples hipótesis. Así, como se

verá en el capítulo de resultados experimentales, los métodos dinámicos permiten obtener salidas ordenadas en función de su probabilidad. Si seleccionamos las N más probables, podremos tratar de forma natural n hipótesis simultáneamente. En la sección 4.4se explicará con bastante detalle cada uno de los métodos utilizados, así como, los parámetros utilizados.

Capítulo 4

Experimentos

En este capítulo se van a comentar los experimentos realizados para analizar la capacidad del modelo BB6-HM para extraer información de la escena con el objetivo de identificar las situaciones y eventos de interés para la tarea de vigilancia descritos en el capítulo 3.

La clasificación de situaciones y eventos de interés se va a realizar empleando dos tipos de métodos: 1) basados en análisis de los parámetros del modelo BB6-HM y aplicando heurísticas y 2) mediante aprendizaje supervisado a partir de datos(ver figura 4.1). Los primeros se utilizarán para identificar situaciones en las que es posible obtener valores o rangos para parámetros del modelo mediante la observación de secuencias y la realización de mediciones en sus fotogramas por parte de humanos. Los segundos se utilizarán en situaciones en las que esos valores o rangos son difíciles de obtener por el método anterior y es preferible entrenar a un clasificador para que se encargue de la labor. Dentro de los clasificadores definidos mediante aprendizaje se han usado clasificadores estáticos y dinámicos, ver sección 2.4. Los clasificadores dinámicos se han usado para identificar situaciones que requerían una secuencia larga de fotogramas y, por tanto, era más acusado su carácter dinámico.

El reconocimiento mediante reglas heurísticas se realiza comprobando las condiciones asociadas a cada situación para el grupo de parámetros que la caracteriza. Este tipo de reconocimiento se ha aplicado a la localización de partes del cuerpo, a la obtención de información sobre el movimiento (velocidad y dirección), al reconocimiento de situaciones y eventos primitivos y a la detección de oclusiones.

El reconocimiento mediante aprendizaje supervisado, estático, se ha aplicado a la determinación de la orientación del individuo con respecto a la cámara y a la detección de la situación de llevar objeto, que también se ha analizado heurísticamente. Para el reconocimiento de personas por la forma de andar (gait) se han empleo tanto clasificadores estáticos como dinámicos, modelos ocultos de Markov.

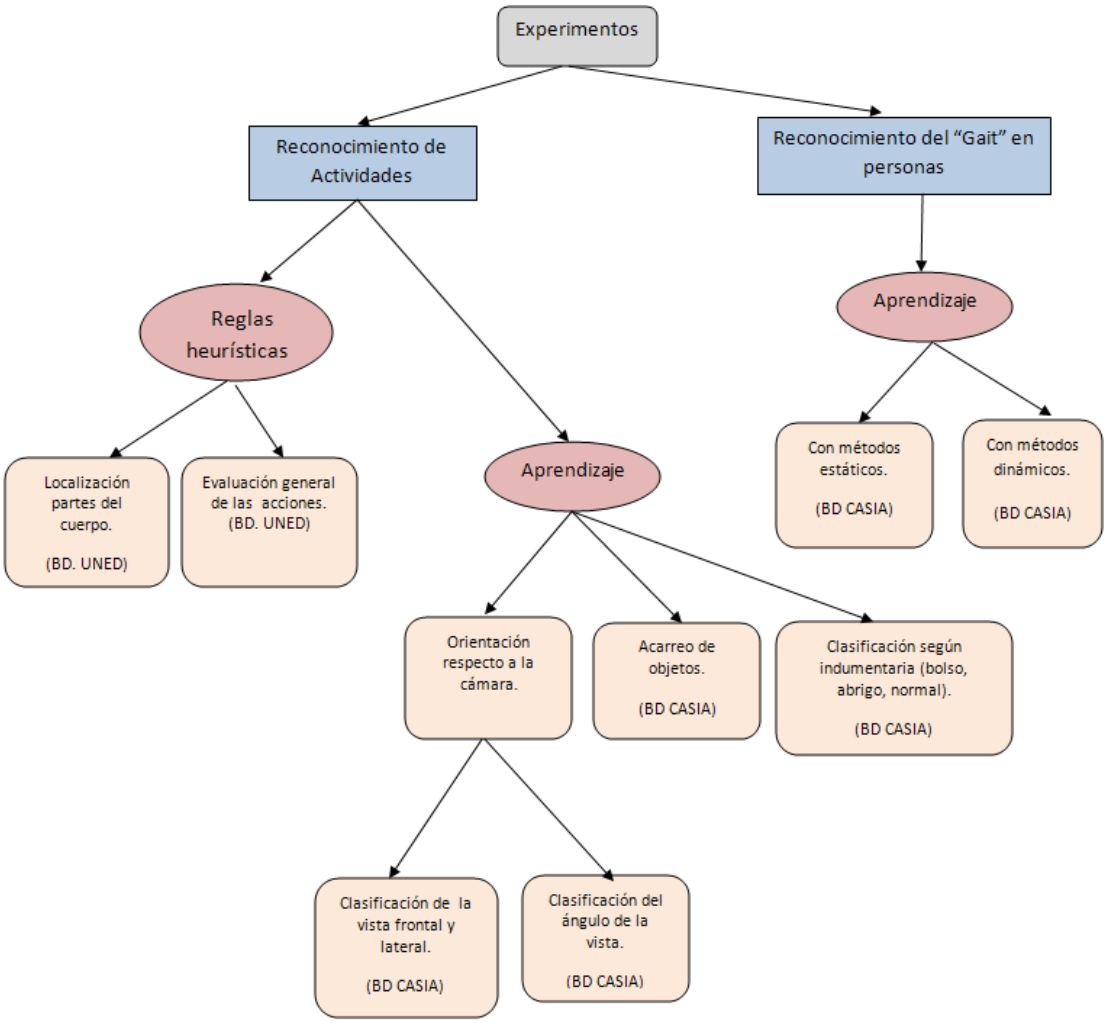


Figura 4.1: Evaluación general de los experimentos realizados dependiendo de las técnicas aplicadas y del tipo de base de datos utilizada.

En las distintas secciones de este capítulo daremos una descripción de las herramientas utilizadas, sección 4.1. A continuación, se describirán en detalle los experimentos realizados usando reglas heurísticas, sección 4.2, los experimentos realizados mediante aprendizaje supervisado, sección 4.3, y los experimentos de reconocimiento del gait, sección 4.4. Terminaremos con un resumen de los resultados de los experimentos, sección 4.5.

4.1. Herramientas utilizadas

4.1.1. Imágenes utilizadas para la evaluación del modelo BB6-HM

Para evaluar la capacidad de reconocimiento del modelo se han tomado las imágenes de diferentes fuentes. Por un lado, imágenes tomadas por nuestro equipo de investigación en las instalaciones de nuestra Universidad (UNED), y por otro, imágenes de la base de datos de CASIA de la Chinese Academy of Sciences.

4.1.1.1. Base de datos de la UNED

Esta base de datos consta de 52 secuencias filmadas a 25 frames por segundo. Las secuencias muestran la aparición de diferentes personas (hombres y mujeres) desde vistas tanto frontales como laterales, que aparecen realizando diferentes acciones: andando lentamente, andando rápidamente, corriendo, levantando los brazos, agachándose, saltando, cogiendo y dejando objetos, etc. Dada la escala de tiempo en la que suceden las acciones realizadas por humanos, resulta que dos frames consecutivas son muy similares, por lo que se eliminó una de cada dos frames para que los cambios entre frames fueran más significativos. A su vez, así se reducía también la carga computacional. La duración de cada una de las secuencias es variable según el tipo de actividad, yendo desde 25 a más de 150 frames después de la selección de una de cada dos, resultando un total de 1505 imágenes. La segmentación de los blobs correspondientes a los humanos en la escena se realizó utilizando el método de los conos truncados desarrollado en la UNED Rincón et al. (2007), el cual es un método basado en la sustracción del fondo que resulta muy robusto a cambios de iluminación y reflejos.

4.1.1.2. Base de datos de CASIA

La base de datos CASIA, Chinese Academy of Sciences (2005) es una gran base de datos multivista que se creó en enero de 2005. Consta de secuencias de 124 personas diferentes, 93 hombres y 31 mujeres, que caminan con distintas orientaciones con respecto a la cámara y con distintas indumentarias. Las orientaciones se distribuyen entre 0° y 180°, con un intervalo de 18°. La indumentaria puede ser: normal (ropa ajustada al cuerpo), con abrigo o portando bolso. Cada persona camina 10 veces en cada una de las escenas, 6 normal, 2 con abrigo y 2 con bolso. En total, hay un conjunto de 13640 secuencias y cada una consta de aproximadamente 90 frames, este número varía entre las diferentes personas, con lo que en total hay aproximadamente

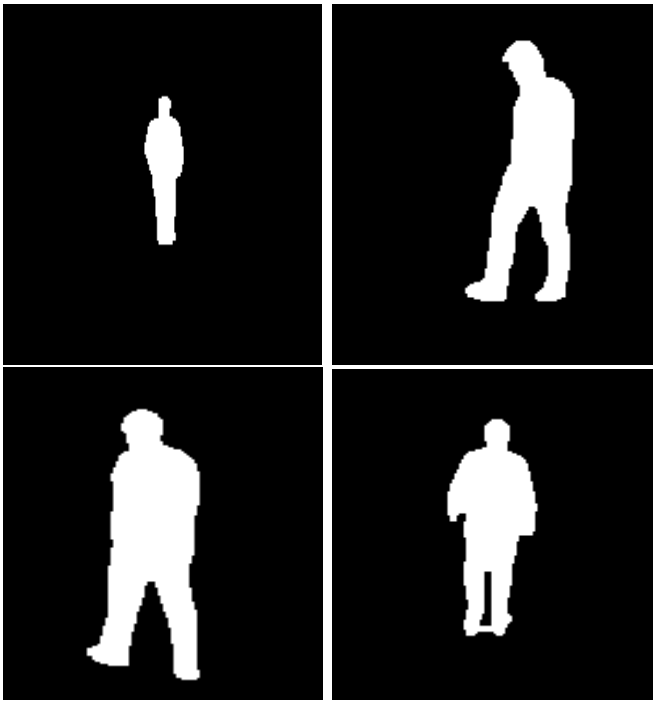


Figura 4.2: Ejemplos de siluetas de la base de datos de CASIA para el ángulo: 0^{o} y 72^{o} para las dos superiores y 144^{o} y 180^{o} para las inferiores.

1.200.000 imágenes.

En nuestros experimentos hemos usado el conjunto de datos Dataset B de CASIA. El formato del nombre de archivo de vídeo en este

conjunto de datos B es: 'xxx-mm-nn-ttt.avi', donde

- xxx: Identificación del individuo, desde 001 hasta 124.
- mm: Estado de pie, puede ser "nm" (normal), "cl" (en un abrigo) o "bg" (con un bolso).
- nn: Número de secuencia.
- ttt: Ángulo de visión, puede ser '000 ', '018', ..., '180 '.

En la fig. 4.2, se muestran unas imágenes de la base de datos de CASIA con diferentes ángulos para la vista. (0^{o}, 72^{o}, 144^{o} y 180^{o}).

La base de datos de CASIA tiene distribuida la información de la siguiente manera, por un lado, cada persona aparece numerada desde el 001-124. Esta numeración corresponde a los directorios, dentro de los cuales, contendrá otros diez directorios cada uno de los cuales aparece denominado según la indumentaria utilizada:

- bg : imágenes con bolso.

- cl : imágenes con abrigo.
- nm : imágenes normales.

Por ejemplo, si tomamos la persona denominada como 13, ésta contendrá un directorio denominado 013 y dentro de éste aparecerán otros diez denominados: bg-01, bg-02, cl-01, cl-02, nm-01, nm-02, nm-03, nm-04, nm-05, nm-06. Los números que aparecen después del guión son simplemente una numeración secuencial por cada tipo. La ruta quedará por tanto como: "....\013\bg-01".

A su vez, dentro de cada uno de estos directorios, aparecen otros once directorios (000, 018, 036, 054, 072, 090, 108, 126, 144, 162 y 180), correspondientes a los diferentes ángulos de la cámara con respecto a la persona.

Para cada directorio, correspondiente a cada ángulo, están todas las imágenes, con el formato siguiente:

- xxx : identificación de la persona.
- mm : situación de la persona (bolso, abrigo, normal).
- nn : número de secuencia.
- ttt : ángulo de la vista $0^{\underline{o}}$, $18^{\underline{o}}$ ... $180^{\underline{o}}$

Esta manera de nombrar las imágenes, hace que su manejo dentro del programa sea mucho más sencillo, pudiendo clasificarlas y separarlas, por persona, ángulos o tipo de situación.

4.1.1.3. Preprocesamiento de imágenes

Previamente a efectuar el tratamiento de las imágenes de la base de datos de CASIA, se ha realizado un análisis del estado de las mismas y se han detectado bastantes imágenes mal segmentadas, por lo que ha sido necesario eliminarlas para facilitar el aprendizaje a partir de datos. Se ha creado un programa, codificado en lenguaje "C", que se encarga de recorrer toda la base de datos y crear el modelo BB6-HM por cada una de las frames. Después analiza cada bloque trazando dentro del rectángulo que lo componen líneas horizontales desde el lado izquierdo al lado derecho y comprobando el número de puntos de corte con el blob. Analizando el número de puntos de corte extrae aquellas imágenes que tienen mala segmentación, entendiéndose por mala segmentación aquellas imágenes que aparecen cortadas de alguna manera. Se ha comprobado que aproximadamente el 10 % de las imágenes se encuentran dañadas. En fig. 4.3 se muestra un ejemplo de humano bien segmentado y dos de segmentación incorrecta. Las imágenes dañadas son eliminadas de la secuencia tras el preprocesado.

Figura 4.3: Imágenes de la base de datos de CASIA. (a) Bien segmentada; (b) Mal segmentada en la cabeza; (c) Mal segmentada en las piernas.

4.1.2. Librerías utilizadas

Para la implementación de los métodos de aprendizaje usados en esta tesis se han utilizado dos librerías: WEKA, para la implementación de modelos mediante aprendizaje supervisado por métodos estáticos, y la toolbox de MATLAB para modelos ocultos de Markov de D. Murphy, para la implementación de modelos mediante aprendizaje supervisado por métodos dinámicos.

4.1.2.1. WEKA

La librería WEKA Group (2012) es una colección de algoritmos de aprendizaje para tareas de minería de datos. Agrupa diferentes herramientas para: preprocesado de datos, agrupamiento o clustering, clasificación, regresión, generación de reglas de asociación, etc. También incluye facilidades para la visualización de los datos. En la investigación desarrollada en esta tesis se han utilizado los algoritmos de J48, Bagging y stacking, perceptrón multicapa (MLP) y máquinas de vectores soporte (SVM), ver sección 2.4.1.

Entre sus características se pueden citar:

- Se trata de uno de los entornos más populares, utilizado en numerosos trabajos de investigación y en múltiples entornos comerciales.
- La interfaz es simple e intuitiva.
- Incluye una de las colecciones más extensas de algoritmos del mercado y permite experimentar con Redes Neuronales, algoritmos basados en reglas o árboles de decisión, modelos bayesianos y probabilísticos como las redes de Inferencia Bayesianas, etc.
- La API de programación es extremadamente simple y con dos o tres llamadas, es posible integrar los algoritmos en el programa.

- La mayoría de las veces, trabajar en Java no es un problema porque en servidor (Servlet's, JSP's) es tan rápido y eficiente como muchos otros lenguajes.

- Es totalmente portable entre sistemas operativos y existen tres versiones principales: para Windows, Mac OS X y Linux.

En el Anexo II, capítulo 8, se explica con detalle la herramienta con las opciones utilizadas en esta tesis.

4.1.2.2. Toolbox para los modelos ocultos de Markov (HMM's)

Para realizar el entrenamiento con HMM's, ver subsección 2.4.2, se utiliza la toolbox HMM de Matlab de Murphy (2005). Esta herramienta permite utilizar HMM tanto en tiempos discretos como continuos.

Para los tiempos discretos se utilizan diferentes funciones según la aplicación. Por un lado, está "dhmm_em" que se utiliza para entrenar un modelo discreto con el algoritmo de EM (realmente utiliza el algoritmo Baum-Welch, que es una simplificación del EM). Otra es la función "dhmm_logprob" que se puede utilizar para clasificar una secuencia dado un modelo, devolviendo como resultado la probabilidad logarítmica de que en un HMM se de un determinado conjunto de observaciones. Esta función permite discriminar entre varios HMM candidatos a generar una determinada secuencia, determinando cuál de ellos es el que tiene más probabilidades de generarla. Y por último, para la obtención de una secuencia de observaciones que incluya datos observados y las transiciones ocultas, se utiliza la función "dhmm_sample". Se puede utilizar para generar secuencias artificiales de pruebas que incluyan, además de la observación, las transiciones ocultas, de forma que facilite los entrenamientos o las evaluaciones de los HMM. En el Anexo III, capítulo **??**, se presentan a modo de resumen las funciones para tiempos discretos y continuos.

En el caso continuo se dispone de funciones similares denominadas: "mhmm_em", "mhmm_logprob" y "mhmm_sample. En este caso, la función "mhmm_em" utiliza funciones gaussianas de densidad de probabilidad para representar las probabilidades de las variables ocultas.

4.2. Experimentos realizados por métodos heurísticos

Estos experimentos consisten en estudiar estudiar el comportamiento de los parámetros del modelo BB6-HM y definir reglas heurísticas para reconocer diferentes

situaciones dadas en las secciones 3.4 a 3.7. El análisis se ejecutó en una ventana temporal de 10 frames utilizando la base de datos de la UNED, la cual se demostró suficiente para determinar los eventos que ocurrían.

Para realizar todos estos experimentos se siguió el siguiente procedimiento, el cual se puede visualizar en la figura 4.4.

- De manera manual se anotan los instantes de tiempo en los que se producen los eventos.
- El programa que analiza las situaciones recorre todas las frames de la secuencia preprocesada evaluando cada uno de los parámetros definidos en el capítulo 3 y produciendo los eventos de manera automática.
- Para evaluar cada evento, se compara, en los diferentes instantes de tiempo, las anotaciones con los resultados obtenidos de manera automática. Hay que tener en cuenta que hay un cierto margen de desfase entre la anotación manual y la automática, por tanto, aunque el evento detectado no coincide exactamente con el evento anotado, se considerará que la detección es correcta porque están muy próximos temporalmente.

4.2.1. Experimentos sobre localización de partes del cuerpo

Uno de los fines del modelo es identificar la posición de las diferentes partes del cuerpo. Tal y como se vio en la subsección 3.4, para localizarlas se analizan los bloques de forma independiente, dependiendo de la parte del cuerpo en la que estemos interesados. La Tabla 4.1 muestra el porcentaje de éxito sobre la localización de estas partes del cuerpo en siete secuencias. Como se puede ver, la aproximación en la localización de la cabeza es muy alta debido al hecho de que en ninguna secuencia el humano elevó sus brazos. En general, los pies se localizan correctamente, pero no ocurre lo mismo con las manos debido a que a veces son ocluidas mientras se camina. Finalmente, la localización del punto de unión de las piernas, P_Λ, es más sensible a las operaciones morfológica realizadas en el proceso de segmentación. A pesar de esto los resultados son bastante satisfactorios. La media del porcentaje de éxito varia entre 82.7 % y 100 %.

4.2.2. Experimentos sobre obtención de información del movimiento

Para la obtención de la información sobre la velocidad del movimiento se han usado los parámetros descritos en la subsección 3.5.1 Los valores para las cotas dadas

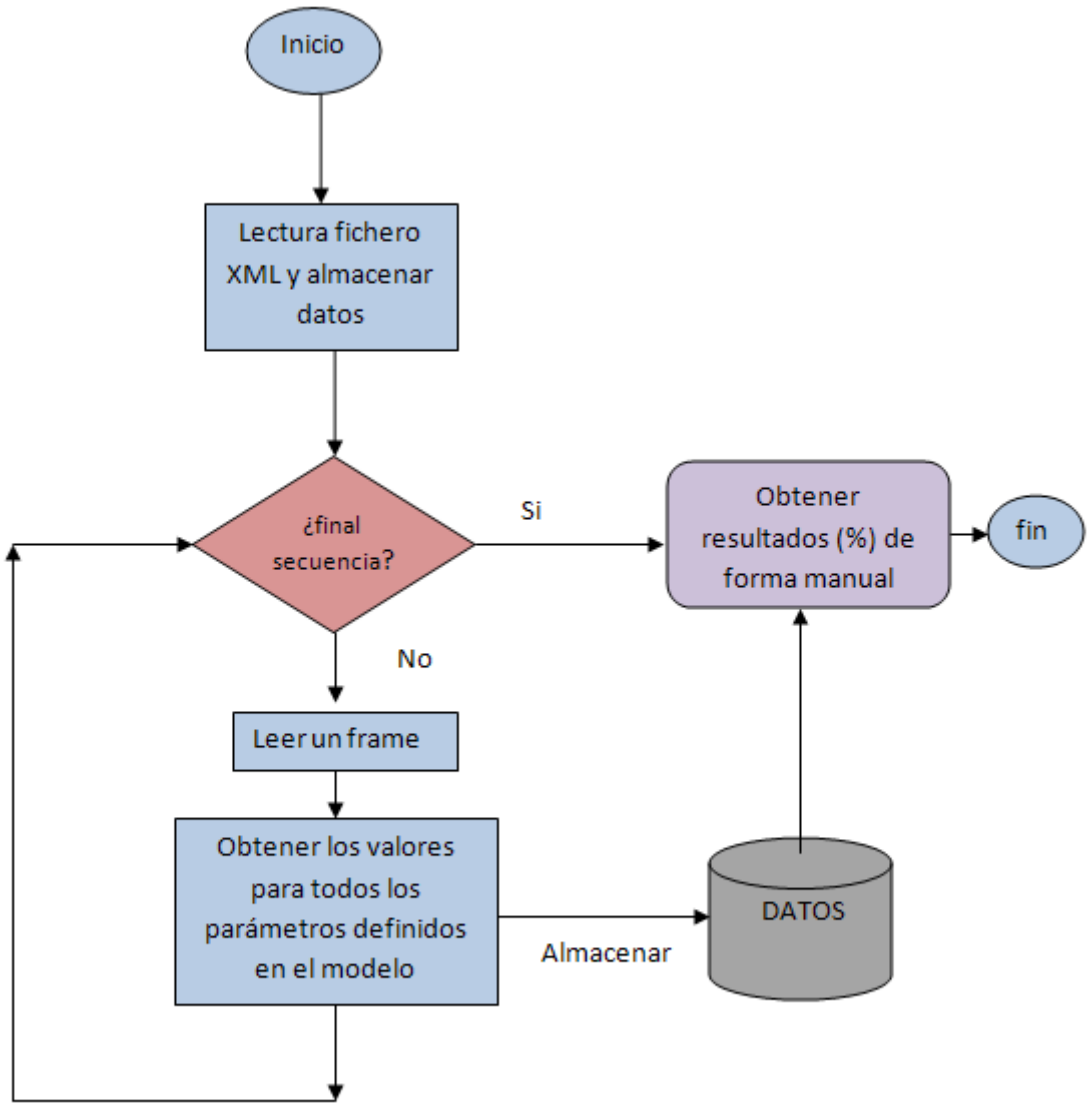


Figura 4.4: Diagrama del proceso para la realización de los experimentos realizados en el caso de utilizar reglas heurísticas. El programa que analiza las situaciones recorre todas las frames evaluando cada uno de los parámetros definidos en el capítulo 3 con la situación que se quiere detectar y sacando por pantalla y en un fichero de salida los resultados de dicho análisis. Por tanto, al final del experimento, se tiene un fichero resultante con el valor de todas las frames analizadas.

Sequence (No. of Frames)	P_{HD} (%)	P_{H1} (%)	P_{H2} (%)	P_{F1} (%)	P_{F2} (%)	$P_{\wedge}$ (%)
1 (90)	95.0	60.0	96.6	100	100	95.0
2 (60)	100	85.0	95.0	100	100	95.0
3 (36)	100	100	77.8	100	100	97.2
4 (73)	100	84.9	60.3	100	100	93.2
5 (73)	100	84.9	91.8	100	100	98.6
6 (100)	100	99.0	99.0	100	96.0	91.0
7 (47)	100	68.1	93.6	100	100	93.6
Average	99.1	82.7	88.9	100	99.2	94.5

Tabla 4.1: Porcentaje de las correctas localizaciones de las partes del cuerpo en diferentes secuencias de vídeo: cabeza, manos, pies y punto de unión de las piernas.

	Lento	Normal	Rápido	Derecha	Izquierda
Lento(122)	121	1			
Normal(372)	2	367	3		
Rápido(36)		3	33		
Derecha(239)				237	2
Izquierda(291)				3	288

Tabla 4.2: Resultados de los experimentos de la obtención de información del movimiento. Cada fila indica un estado o evento y cada columna su clasificación. El número de veces que éste estado o evento ocurre en realidad aparece entre paréntesis en la primera columna.

en esa subsección para la distinción entre movimiento lento, normal o rápido han sido:

$$\text{DIST\_LENTO} = \frac{H_T}{60}$$

$$\text{DIST\_RAPIDO} = \frac{H_T}{20}$$

$$\theta_s = 10^\circ$$

$$\theta_f = 25^\circ$$

Además de la velocidad también se han realizado experimentos sobre la información de la dirección del movimiento, derecha o izquierda.

Los resultados se muestran en la tabla 4.2. Como puede observarse, el porcentaje de acierto es muy elevado, el peor caso se da al intentar reconocer un movimiento rápido siendo el porcentaje de acierto del 92 %.

	NAO	AO	BL	G	AG	LV	RO	DO	S	ND
NAO(25)	24	1								
AO(18)	2	16								
BL(6)			4	2						
G(28)				23						5
AG(3)					3					
LV(3)						3				
RO(2)							2			
DO(2)								2		
S(3)						1			2	

Tabla 4.3: Resultados de los experimentos sobre el reconocimiento de situaciones primitivas. Cada fila indica un estado o evento y cada columna su clasificación. El número de veces que este estado o evento ocurre en realidad aparece entre paréntesis en la primera columna. NAO: No Acarrea Objeto; AO: Acarrea Objeto; BL: Brazos levantados; G: Giros; AG; Agacharse; LV; Levantarse; RO: Recoger un Objeto; DO: Dejar un Objeto; S: Saltar; ND: No Detectado.

4.2.3. Experimentos sobre reconocimiento de situaciones y eventos primitivos

Las situaciones y eventos primitivos que se han intentado reconocer en este experimento son los descritos en la sección 3.6 usando, en cada caso, las condiciones expresadas para cada evento o situación en dicha sección. La lista completa puede verse al pie de la tabla 4.3.

Los resultados se muestran también la tabla 4.3 Como puede observarse, el porcentaje de acierto es muy elevado, el peor caso se da al intentar reconocer las acciones de brazos levantados y de girar con unos porcentajes de acierto del 66 %. Los resultados son muy prometedores, ya que en bastantes casos se consiguen porcentajes de acierto del 100 %, aunque en muchos de ellos el número de situaciones existentes es demasiado pequeño para extraer una conclusión estadísticamente significativa..

4.2.4. Experimentos sobre detección de oclusiones

Para el caso de la oclusión se eligieron algunas frames de manera aleatoria y se añadió un objeto rectangular ficticio definido por sus puntos extremos, tal y como se comentó en **??**. Estas coordenadas se introducen en un fichero XML para posteriormente ser leído por el programa que se encarga de analizar la situación. Como el obstáculo es estático, sus coordenadas no se introducen por cada frame, si no que se almacenan en un registro de objetos que pueden ocluir y que está asociado a la escena. Al final, de manera manual, se procede a verificar el porcentaje de

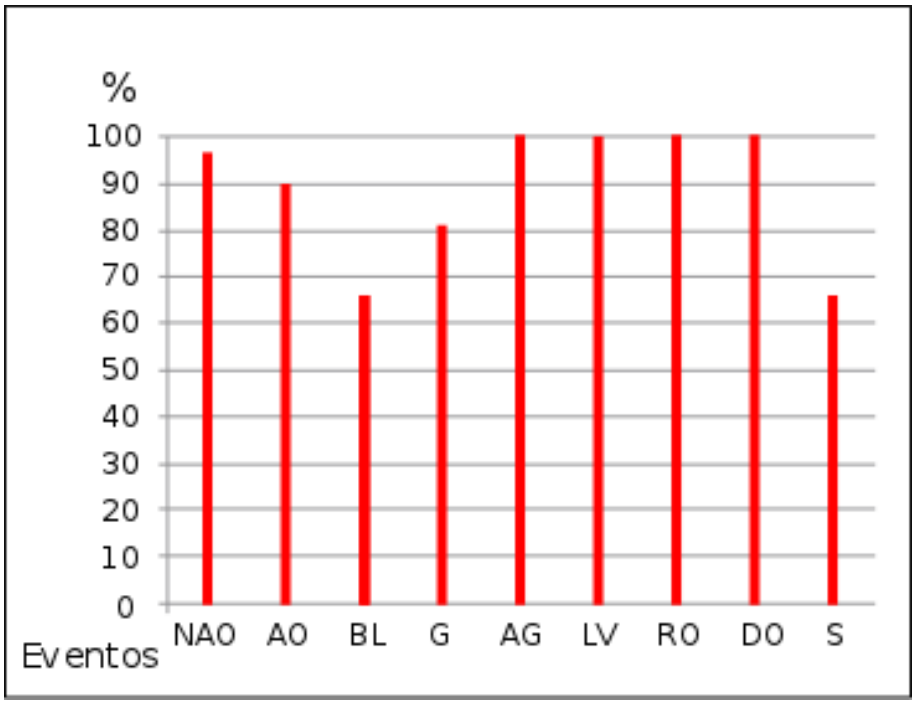


Figura 4.5: Porcentaje de éxito por eventos analizados. Se muestra el porcentaje de exactitud para cada tipo de evento. Revisando los valores de esta gráfica, se observa que el porcentaje de aciertos en la detección de determinadas situaciones es bastante alto. Por ejemplo, en los casos de oclusiones totales, agacharse, levantarse, recoger un objeto o dejar un objeto, tienen un acierto del 100 %. Sin embargo en el caso de la detección del evento "saltar", el valor de la gráfica cae hasta un 66 %. Para la nomenclatura ver pie de tabla 4.3

	Oclusión parcial	Oclusión total	ND
Oclusión parcial(65)	63	0	2
Oclusión total(16)	0	16	0

Tabla 4.4: Resultados de los experimentos de la detección de oclusiones. Cada fila indica un estado o evento y cada columna su clasificación. El número de veces que éste estado o evento ocurre en realidad aparece entre paréntesis en la primera columna. ND indica no detectado.

aciertos.

En la detección de oclusiones producidas por obstáculos fijos rectangulares se han obtenido elevados porcentajes de acierto del 97.5 % para la oclusión parcial y del 99 % para la total como puede apreciarse en la tabla 4.4.

4.3. Experimentos realizados mediante aprendizaje supervisado

En el caso de aprendizaje supervisado se ha utilizado la herramienta de WEKA y, principalmente, la base de datos de CASIA, aunque en alguno casos se usó la de la UNED. Se ha creado un programa en "C" al que se le ha pasado como entrada el conjunto de secuencias de la base de datos y se ha obtenido su descripción en base al modelo BB6-HM. Esta descripción se ha grabado en un fichero de formato WEKA (arff) para después proceder al entrenamiento de los clasificadores y a su evaluación.

Cada uno de los métodos en WEKA, dispone de un conjunto muy amplio de métodos y opciones para evaluar los diferentes conjuntos de muestras. De estos, se han elegido aquellos que proporcionaban mejores resultados, que se muestran en la Tabla 4.5. Para la división del conjunto de secuencias en los subconjuntos de entrenamiento y test se ha utilizado diversos métodos:

1. Validación cruzada o cross-validation, que es una técnica utilizada para evaluar los resultados de un análisis estadístico y garantizar que son independientes de la partición entre datos de entrenamiento y prueba. Consiste en repetir y calcular la media aritmética obtenida de las medidas de evaluación sobre diferentes particiones. Se utiliza en entornos donde el objetivo principal es la predicción y se quiere estimar cómo de preciso es un modelo. El valor en las pruebas ha sido de 10 iteraciones o 10-fold-validation, donde los datos de la muestra se dividen en 10 subconjuntos, de los cuales uno se utiliza como test y el resto como datos de entrenamiento.

2. Porcentage-split en el se dividen los datos en dos grupos, de acuerdo con el

MÉTODOS	PARÁMETROS UTILIZADOS		VALORES
J48	Factor de Confianza	-C	0.25
	Número mínimo de instancias por hoja.	-M	2
AdaBoost	Semilla a partir de la cuál se generan los números aleatorios del algoritmo.	-S	1
	Número de iteraciones del algoritmo si no converge antes.	-I	10
	Umbral para realizar la poda	-P	100
	Clasisificador seleccionado	-W	J48
Bagging	Semilla a partir de la cuál se generan los número aleatorios del algoritmo.	-S	1
	Número de iteraciones del algoritmo si no converge antes.	-I	10
	Umbral para realizar la poda	-P	100
	Clasisificador seleccionado	-W	REPTree
Stacking	número de subconjuntos en que hay que dividir el conjunto de ejemplos para, el último de ellos, emplearlo como conjunto de test	-X	10
	Clasificador selacionado	-M	J48
	Semilla a partir de la cuál se generan los números aleatorios del algoritmo.	-S	1
	Meta-Clasificador selecionado	-B	J48
MLP	el número de neuronas en la capa oculta	-H	5
	Tasa de aprendizaje	-L	0.3
	Tiempo de Entrenamiento	-N	1000
	momento	-M	0.2
SVM	Valor de Epsilon	-P	1.0E-12
	Semilla para generar los números aleatorios que inicializarán los parámetros	-W	1
	Parámetro de Tolerancia	-L	0.0010
	Kernel	-K	PolyKernel

Tabla 4.5: Tabla con los comandos de ejecución para cada uno de los métodos evaluados con las opciones que han resultado más óptimas en los experimentos.

porcentaje indicado (%). El valor indicado es el porcentaje de instancias par construir el modelo, que seguidamente es evaluado sobre las que se han dejado aparte. El valor utilizado en esta tesis ha sido del 66 %.

3. Fichero de datos de test se eligió de dos maneras por un lado tomando el 10 % de imágenes y por otro lado una secuencia completa del conjunto global de imágenes de la persona a reconocer.Este porcentaje y la secuencia se eligieron de manera aleatoria.

4.3.1. Evaluación de la orientación con respecto a la cámara: vistas frontal y lateral y ángulo de la vista.

Dos experimentos diferentes se realizan para evaluar la orientación del humano con respecto a la cámara, que recordemos varía de 0° a 180° en intervalos de 18° en las secuencias de CASIA, ver subsección 4.1.1. Por un lado, se ha generado un modelo para distinguir entre vista lateral y frontal, ya que el significado de los parámetros del modelo para la detección de varios eventos depende bastante de ello. Por otro lado, se obtiene otro modelo para determinar el ángulo aproximado formado con respecto a la cámara. En ambos casos se ha utilizado WEKA para su evaluación.

1. Reconocimiento de las vistas lateral y frontal

   Para la clasificación de este caso se han utilizado los métodos de J48 , Meta-clasificadores, MLP y SVM haciendo corresponder las orientaciones 0º y 180º con la clase vista frontal y el resto con la lateral. Para lo cual se creó un nuevo parámetro de anotación denominado Vista_lateral_frontal", cuyos posibles valores son FRONTAL (F) o LATERAL (L).

   Partiendo de los parámetros ya definidos en 3 se han realizado un conjunto de experimentos para determinar los más discriminantes. Los mejores resultados se han obtenido utilizando las anchuras de los bloques B_2, B_5 y B_6 (W_2, W_5, W_6), obteniéndose un porcentaje de aciertos del 99,5 % con el método de SVM, sobre el conjunto de test. Este resultado demuestra que la información se puede extraer de un número pequeño de parámetros.

   Por otra parte, los errores suelen aparecer en imágenes que tienen una mala segmentación, lo cual hace incorrecta la información asociada a varios parámetros del modelo.

2. Reconocimiento del ángulo del humano respecto a la cámara

```
=== Confusion Matrix ===
    a    b    c    d    e    f    g    h    i    j    k   <-- classified as
 3509   87   21    0    0    0    0    0    3    6  237 |    a = 0
   88 3012  344   59    1    0    1   20   74  175   65 |    b = 18
   24  362 2757  326   36   13   12   44   80  144   17 |    c = 36
    0   47  374 2890  158   69   86  124  100   36    0 |    d = 54
    0    2   34  147 2905  427  225  114   49    8    0 |    e = 72
    0    2   21   73  484 2861  358   62   12    0    0 |    f = 90
    0    1   26   75  255  405 2824  262   62    2    0 |    g = 108
    0   14   43  129  117   85  247 2751  335   36    0 |    h = 126
    3   53   59   85   39   25   91  376 2966  251    5 |    i = 144
    6  159  132   31   14    1    6   55  233 3179   26 |    j = 162
  274   38   13    0    0    0    0    0    9   24 3461 |    k = 180
```

Figura 4.6: Matriz de confusión obtenida con WEKA, del reconocimiento del ángulo de la persona respecto a la cámara. En dicha matriz, la matriz diagonal indica el número de instancias clasificadas correctamente a partir del conjunto de "test". En cuanto a los errores, se aprecia que la mayor parte del error se clasifica como los ángulos vecinos o el inverso (las primeras diagonales abajo y encima diagonal principal).

Al igual que en el caso anterior, partiendo de los parámetros proporcionados en 3, usando el método J48, Metaclasificadores, MLP y SVM se han obtenido los parámetros más discriminantes. Los mejores resultados se han obtenido para el vector de características dado por:

$$\left(W_2, W_5, W_6, DS_2, DS_6, \Delta CM_M^6\right)$$

El porcentaje que se obtuvo en la clasificación fue del 94,75 % con el método de SVM sobre el conjunto de test. Estos resultados vuelve a confirmar que la información puede extraerse de un número pequeño de parámetros.

En fig.4.6, se muestra la matriz de confusión para el reconocimiento del ángulo que forma una persona respecto a la cámara utilizando el método J48. En dicha matriz, la matriz diagonal indica el número de instancias clasificadas correctamente a partir del conjunto de "test". En cuanto a los errores, se aprecia que la mayor parte del error se clasifica como los ángulos vecinos o el inverso (las primeras diagonales abajo y encima diagonal principal), pues en la silueta no hay información de si el blob mira hacia adelante o hacia atrás. En algunas aplicaciones, es suficiente indicar la orientación aproximada, por lo tanto, si consideramos como correcto el ángulo con una desviación de $\pm 18^{\text{o}}$, la tasa de éxito (la suma de la diagonal principal y adyacentes) es 90.5 %. Además, como en caso anterior, debemos tener en cuenta los errores de segmentación.

4.3.2. Reconocimiento del evento llevar objeto

En este caso, se evalúa la posibilidad de los humanos llevando objetos (ver 3.6.10). Inicialmente, el parámetro HC se utiliza al ser uno de los parámetros más significativos para determinar este evento desde el punto de vista lateral. En la fig.4.7 se analiza la influencia de este parámetro para la detección del evento de llevar objeto. En esta figura, se muestra la evolución temporal del HC en el caso de llevar y no llevar objeto. Hay que destacar cómo la evolución temporal es periódica en ambos casos, pero la fundamental diferencia es este caso la amplitud, la cual resulta mucho mayor cuando el humano no está llevando objeto. Así mismo, la solución a este caso no tiene por qué estar reducida al estudio de este único parámetro, ya que, como se ha comentado en diversas ocasiones, no es suficiente para el caso de vista frontal o cuando el humano lleva algún tipo de objeto de dimensiones pequeñas.

Para modelar este caso se ha utilizado el método J48 y se han definido dos clases: no llevando objeto y llevando objeto. Partiendo de los parámetros del modelo, se han realizado varios experimentos que han llevado a determinar los parámetros más discriminantes. Los mejores resultados, con un porcentaje del 86 %, se han obtenido para el vector de características:

$$(HC, W_2, W_3, W_5, W_6, S_2, S_4, S_6, S_f, S_h)$$

La justificación de este vector de características está en el hecho de que en el reconocimiento de llevar objeto los bloques implicados son los bloques de en medio y es por eso que los parámetros de este vector son los relacionados con estos bloques. Además, en vista lateral el parámetro HC, es uno de los más discriminantes.

4.3.3. Reconocimiento de la apariencia según la indumentaria: Normal (NM), Lleva Bolso (BG), Lleva Abrigo (CL)

En este caso, se trata de clasificar los tres tipos de casos que vienen en la base de datos de CASIA. Es decir, se trata de analizar a qué grupo de los tres (normal (NM), lleva bolso (BG), lleva abrigo (CL)) pertenece el individuo presente en una escena. Para ello se han definido tres clases nombradas como: normal, lleva bolso y lleva abrigo. Se han utilizado los métodos J48, Metaclasificadores, MLP y SVM. Partiendo de los parámetros del modelo se han realizado varios experimentos que han llevado a determinar los parámetros más discriminantes. Los mejores resultados se han obtenido para SVM con un resultado de 90,2 % y para el vector de caracte-

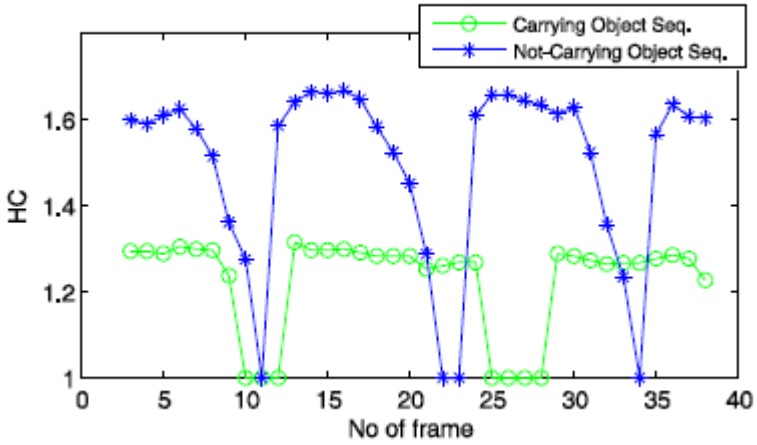


Figura 4.7: . Evolución temporal del parámetro HC desde el punto de vista lateral: caso llevando y no llevando bolso.

rísticas: .

$$\left(\theta, \alpha, \Delta CM_M^5, H_4, H_5, W_4, W_2, W_3, W_6, S_4, HC, S_f, S_h S_{B4\_B5}\right)$$

En este caso, la justificación para este vector obtenido viene dada por el hecho de que si la persona lleva bolso los bloques más afectados son los de en medio. En el caso de llevar abrigo afecta a bloques más inferiores. En el caso normal, la variación de parámetros involucra a un conjunto más amplio de parámetros, ángulos bloques medios e inferiores.

4.4. Reconocimiento del "Gait" de personas

En este apartado se comentan los experimentos realizados para obtener información sobre el "Gait" o forma de caminar de las personas tal y se comentó en la subsección 3.8. Para realizar los experimentos se ha tomado la base de datos de CASIA y se ha procedido a pasar como entrada a un programa el conjunto de secuencias de esta base de datos.

Como se ya comentó en **??**, la base de datos de CASIA consta de diferentes tipos de situaciones dentro del conjunto de imágenes (con bolso, normal, con abrigo). La evaluación, al igual que se hizo en el reconocimiento de situaciones, se realiza con todas y cada una de estas situaciones, así como, con todas y cada una de las secuencias de las 124 personas.

Para realizar la evaluación de los experimentos tienen dos fases bien diferenciadas: la de entrenamiento y la de test. Tanto para la fase de entrenamiento como

para la de test se extraen, de cada una de las secuencias de imágenes, una de cada cinco imágenes consecutivas como en el caso del reconocimiento de situaciones. El entrenamiento se ha realizado sobre el conjunto de datos a excepción de los utilizados para el test. El conjunto de test se ha obtenido en el caso estático mediante cross-validation, porcentage-split y fichero de test (ver sección **??**) y en el dinámico sólo fichero de test. Para crear el fichero de test, como las frames son muy parecidas, es por ello se han tomado una de cada cinco, pero aún así, van a estar muy próximas a las que se utilizan para entrenamiento, es por esta razón por la que se toma una secuencia completa, de manera aleatoria, de una persona cualquiera y que no se usa en el entrenamiento.

4.4.1. Reconocimiento del "Gait" de personas mediante modelos estáticos

Para la realización de los experimentos se ha utilizado el mismo programa que se utilizó en la sección 4.3. Una vez que el programa ha extraído toda la información de los parámetros del modelo de bloques BB6-HM, se guarda en ficheros de texto con el formato que utiliza WEKA (arff). Esto se hace para todas y cada una de las frames de la secuencia con la que se va a experimentar.

Para realizar el aprendizaje, en caso de probar con fichero de test, se han obtenido dos grupos, por un lado se han separado el grupo de entrenamiento y por otro el de prueba. Una vez se tiene el fichero de WEKA con todos los datos informados, se procede a realizar el test. Un vez se tienen estos ficheros se procede a realizar el entrenamiento utilizando WEKA, en una primera fase se entrena con el fichero de entrenamiento obteniendo un modelo que luego se aplica al test.

Como clases de salida para el entrenamiento se utiliza el identificador de la persona (1..124) proporcionado por la base de datos de CASIA. Partiendo de los parámetros del modelo se han realizado varios experimentos que han llevado a determinar los parámetros más discriminantes.

Los mejores resultados se han obtenido para el vector de características:

$$\left(\theta, \alpha, \Delta CM_M^5, H_4, H_5, H_3, W_4, W_2, W_3, W_5, W_6, S_2, S_4, S_6, S_f, S_h, S_{B4\_B5}\right).$$

En el reconocimiento del gait tiene mucho que ver el movimiento de las extremidades, es por tanto que los bloques involucrados van a ser casi todos. A excepción del bloque correspondiente a la cabeza, el resto de bloques que tienen que ver con las extremidades aparecen involucrados.

Los tiempos de cómputo en el entrenamiento, para el reconocimiento de una persona dada y considerando toda la base de datos completa a excepción de los datos utilizados para las pruebas de test, mediante árboles de decisión (J48) fue de aproximadamente 6-8 horas, mientras que utilizando SVM y MLP se llego a tardar entre 14-18 horas. El número de atributos (variables obtenidas del modelo) utilizados fue inicialmente de 119, un número bastante elevado, lo que hacía que los tiempos se disparasen, por esta razón se aplicaron métodos de reducción de variables, tal y como, el análisis de componentes principales (PCA) para reducir la información redundante y así disminuir los tiempos de cómputo. De esta manera se consiguió reducir el número de atributos a 17 obteniéndose incluso mejores resultados que con los 119 iniciales. Después de aplicar dichos métodos y reducir el número de variables, los tiempos pasaron a reducirse a la mitad. En la sección 4.5 se muestra una tabla resumen con tiempos obtenidos.

Para realizar las pruebas se aplicó porcentage-split ya que sirvió para ir obteniendo una información aproximada de qué resultados se obtenían al aplicar los parámetros que se iban seleccionando y que este método es mucho más rápido a la hora de realizar la evaluación que el cross-validation. Sin embargo una vez estimados los parámetros definitivos que se eligieron en cada momento el método que se tomo para elaborar los resultados fue el cross-validation. Al igual que en el caso del percentage-split, la elección de los ficheros de test ayudaron a realizar las pruebas de manera más rápida que el cross-validation y por tanto extraer información relativa a los parámetros definitivos. . A continuación, en la Tabla 4.6, se muestran los resultados obtenidos mediante cross-validation, correspondientes a las personas etiquetadas como 12, 85 y 15, las cuales se seleccionaron de manera aleatoria. Dichos resultados dependen del método aplicado, la situación y el ángulo con respecto a la cámara. . Al igual que en el caso anterior, se obtienen buenos resultados para las SVM pero se aprecia que en el caso de que la persona lleve abrigo los resultados son peores. Esto se debe a que muchos de los parámetros del modelo sufren menos variación al llevar el abrigo, en especial los que dependen de la variación de la apertura de las piernas. Comentar que los árboles de decisión (J48), en general, producen unos resultados peores que otros métodos, pero son de gran ayuda en las primeras fases del análisis, ya que permiten evaluar la importancia y jerarquía de las variables.

Este análisis se aplica a todas las personas con todas las secuencias existentes, obteniendo como resultado un conjunto de datos que se han plasmado en la fig. 4.8. En esta gráfica los datos se agrupan según el tipo de algoritmo utilizado (J48, SVM, HMM... etc) y el tipo de situación de la imagen (normal, bolso, abrigo), en el caso de los metaclasificadores, para reducir la representación gráfica y puesto que

Persona/Ángulo	Situación	Método	Salida	Persona/Ángulo	Situación	Método	Salida	Persona/Ángulo	Situación	Método	Salida
12/0º	Normal	J48	91.8919 %	85/90º	Normal	J48	93.75%	15/180º	Abrigo	J48	72.15%
		AdaBoost	92.233%			AdaBoost	94.224%			AdaBoost	75,32%
		Stacking	91.323%			Stacking	94.33%			Stacking	76.54%
		Bagging	93.989%			Bagging	96.93%			Bagging	78.94%
		MLP	92.4444%			MLP	98%			MLP	76.88%
		SVM	92.989%			SVM	99.01%			SVM	83.76%
	Bolso	J48	88.888%		Bolso	J48	91.5%		Normal	J48	90.777%
		AdaBoost	91.233%			AdaBoost	92.434%			AdaBoost	91.323%
		Stacking	92.085%			Stacking	93.489%			Stacking	92.323%
		Bagging	92.12%			Bagging	94.99%			Bagging	92.732%
		MLP	100%			MLP	92.23%			MLP	91%
		SVM	100%			SVM	92.819%			SVM	91.898%
	Abrigo	J48	78.9 %		Abrigo	J48	82.5%		Bolso	J48	86.45%
		AdaBoost	81.449%			AdaBoost	82.89%			AdaBoost	87.12%
		Stacking	80.233%			Stacking	80.32%			Stacking	88.24%
		Bagging	82.323%			Bagging	82.23%			Bagging	86.665%
		MLP	82.8988%			MLP	82.898%			MLP	87.376%
		SVM	83.676%			SVM	83.17%			SVM	88.19%

Tabla 4.6: Tabla con los resultados de la evaluación de las personas 12, 15, 85 con diferentes ángulos, situación y métodos. Se obtienen buenos resultados para las SVM pero se aprecia que en el caso de que la persona lleve abrigo los resultados son peores.

los valores obtenidos son muy similares, han agrupado en una variable META que es el promedio de los tres. Puesto que el número de individuos es alto (124), los valores mostrados respecto al porcentaje de aciertos es el valor medio.

En el análisis de cada uno de los casos se ha obtenido unos valores que se han plasmado en la fig. 4.8. En el eje de abscisas se situará el ángulo de la cámara y en el de ordenadas el porcentaje de aciertos correspondientes a dicho ángulo. Por cada uno de los ángulos se representan las barras a diferentes colores con cada uno de los algoritmos de aprendizaje utilizados, así como, la situación en la que se encuentra (normal,bolso o abrigo). De esta manera se podrá ver la diferencia del porcentaje de aciertos por ángulo en función de cada uno de los algoritmos utilizados.

En los experimentos realizados se obtiene que la media de los mejores valores alcanzados es para ángulos entre 72^{o} y 90^{o} en la mayor parte de los casos (vista lateral). Los peores resultados se obtienen para la vista frontal con 180^{o} o valores próximos.

A modo de resumen, se muestra la Tabla **??** donde se recoge un resumen de la fig. 4.8. La tabla se agrupa según el tipo de algoritmo utilizado (J48, SVM, etc) y el tipo de situación de la imagen; N: normal, B: bolso, A: abrigo. La columna “Max Valores”, corresponde los valores máximos del ángulo de la cámara (grados) y “ % resultados” será los valores obtenidos (en porcentaje), para esos ángulos. De igual modo, la columna “Min Valores” corresponde los valores mínimos del ángulo de la cámara (grados) y “ % resultados” será el valor obtenido (en porcentaje) para esos ángulos también.

Observando dicha tabla se aprecia que, en general, los mejores resultados se obtienen para las Máquinas de Vectores de Soporte (SVM) y los peores para los Árboles de Decisión. Al igual que en el caso de reconocimiento de situaciones, el uso de estos últimos sirvió para dar una visión global de las variables más relevantes a considerar en la evaluación.

4.4.2. Reconocimiento del “Gait” de personas mediante modelos dinámicos (modelos ocultos de Markov)

En este caso se procede a realizar el reconocimiento de personas en lugar de situaciones como en apartados anteriores. Para este conjunto de experimentos se ha utilizado la base de datos de CASIA y el toolkit de Murphy (**??**) que trabaja sobre Matlab (versión R2008a). Para realizar al análisis, primero se ha construido un programa interface en lenguaje “C”, que se encarga de pasar los parámetros obtenidos

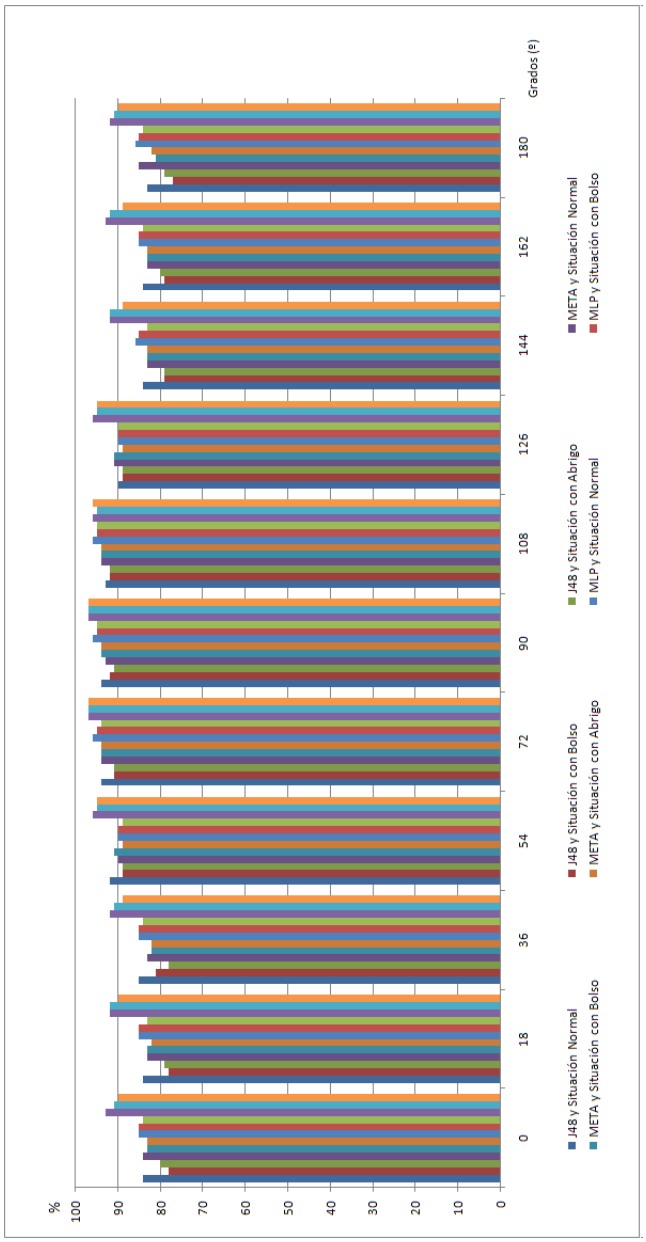


Figura 4.8: Resultados de los experimentos mediante aprendizaje supervisado según diferentes clasificadores y situaciones (normal, bolso o abrigo). En el eje de abscisas se situará el ángulo de la cámara y en el de ordenadas el porcentaje de aciertos correspondientes a dicho ángulo. Por cada uno de los ángulos se representan las barras a diferentes colores con cada uno de los algoritmos de aprendizaje utilizados, así como la situación en la que se encuentra (normal,bolso o abrigo).

del modelo bloques a Matlab. Este programa, al igual que en casos anteriores, se ha encargado de leer todas las secuencias de la base de datos de CASIA e ir obteniendo todos los parámetros comentados en el modelo. Matlab es una herramienta con un gran consumo de memoria, por lo que en diversas pruebas, al ser CASIA una base de datos tan grande, el programa se queda "colgado". Por esta razón, se procedió a partir los datos para tratarlos en Matlab en pequeños paquetes. Esto supuso un trabajo bastante engorroso, ya que había que hacer un preprocesamiento de los datos obteniendo paquetes más pequeños e ir almacenandonos y borrándolos a medida que se utilizaban con el fin de que no se quedase sin memoria. Los tiempos de entrenamiento oscilaron entre 5-8 horas con esta fragmentación de los datos (sección 4.5). En caso de no hacer este preprocesamiento los tiempos llegaban a ser del doble.

Las pruebas consisten en realizar una serie de entrenamientos de la persona en cuestión e ir pasándole después el fichero de test . Para crear el fichero de test, como las frames son muy parecidas, es por ello se han tomado una de cada cinco, pero aún así, van a estar muy próximas a las que se utilizan para entrenamiento, es por esta razón por la que se toma una secuencia completa, elegida de manera aleatoria, de una persona cualquiera y que no se usa en el entrenamiento. Los resultados dan una salida de probabilidades ordenadas de mayor a menor valor. Se escogen las cinco primeras probabilidades ya que se vio que en la mayor parte de los casos la persona elegida quedaba con una probabilidad contenida entre estas cinco.

Al igual que en el caso anterior, se utiliza una única clase: persona que varia entre 1 y 124, correspondiente a cada persona y como vector de características:

$$\left(DS_6, DS_4, \theta, \alpha, \Delta CM_M^5, H_4, H_5, W_4, W_2, W_3, W_5, W_6, S_2, S_4, S_6, S_f, S_h S_{B4\_B5}\right)$$

En el reconocimiento del gait, al igual que pasaba en los estáticos, tiene mucho que ver el movimiento de las extremidades, es por tanto que los bloques involucrados van a ser casi todos. A excepción del bloque correspondiente a la cabeza, el resto de bloques que tienen que ver con las extremidades (brazos, piernas o píes) aparecen involucrados.

Para los modelos de Markov se han realizado pruebas con diferentes números de estados (Q) e iteraciones (It), esto es, que una vez inicializado un HMM éste se ejecuta pasando por los diferentes estados (Q), hasta que se cumplen las condiciones de convergencia o hasta que se alcanza un número máximo de iteracciones (It). Los valores de "It" han variado entre, 1 y 10, encontrándose que los mejores resultados se han obtenido para It=5, no mejorando al aumentar el número de iteraciones pero si la carga computacional. En las pruebas realizadas, también se ha obtenido que con el número de estados Q=4 se han obtenido los mejores resultados. El cambio

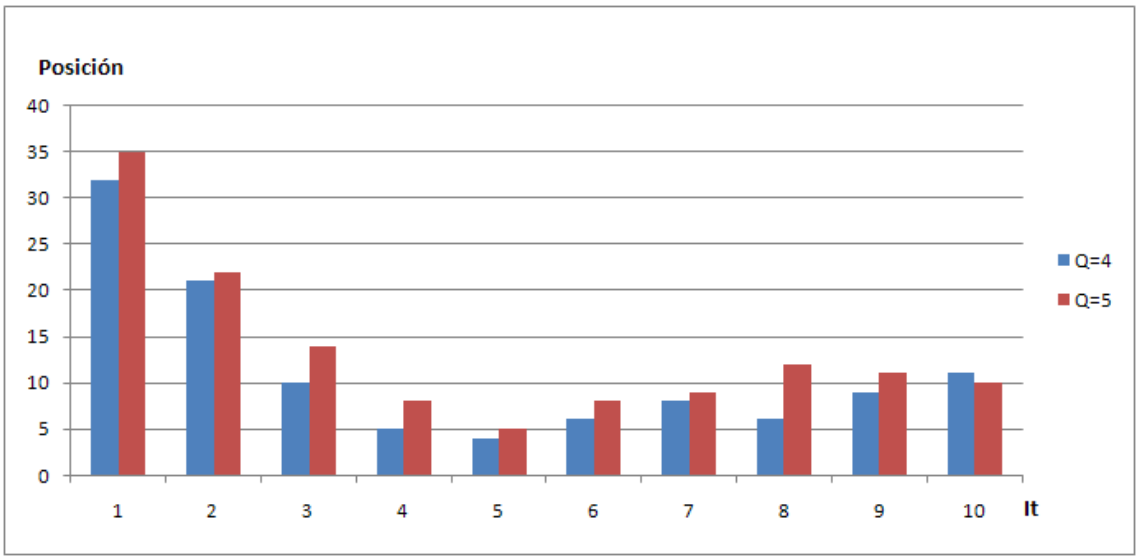


Figura 4.9: Resultados medios de reconocimiento del "Gait" en función de los parámetros Q e It. Se muestra la media de las posiciones en la que se detecta al individuo correcto para el caso de It desde 1 hasta 10 y para los estados 4 y 5 para el conjunto de todas las pruebas realizadas. Los mejores resultados (el individuo correcto está entre los N más probables) se han obtenido para It=5 y Q=4, resultando N=4.

del número de estados, como mucho ha igualado el resultado con Q=4 pero no lo ha mejorado. En la fig.4.9, se muestra la media de las posiciones alcanzadas (mayores probabilidades) para el caso de It desde 1 hasta 10 y para los estados 4 y 5 para el conjunto de todas las pruebas realizadas. En ella se aprecia lo comentado, que los mejores resultados (menor posición) se han obtenido para It=5 y que, en este caso, con Q=4 se obtienen mejores valores que con Q=5.

El valor para el tipo de matriz de covarianza (Cov_type) se ha fijado en 'full' o completa que es el que ha obtenido mejores resultados. De igual manera, para el parámetro número de mezclas gaussianas en la distribución de probabilidad de observación (M), en el que se han obtenido los mejores resultados ha sido para un valor de M=4.

Así, teniendo en cuenta la toolbox de Murphy, ya comentada anteriormente, y, utilizada en esta tesis, los parámetros con los que se han obtenido los resultados serían:

1. Q=4 : nº de estados del modelo.

2. M=4 : nº de mezclas gaussianas en la distribución de probabilidad observación.

3. It=5 : nº de iteraciones.

4. Cov_type='full' :Tipo de matriz de covarianza, puede ser de tipo diagonal

('diag'), completa ('full') o esférica ('spherical'). En esta tesis se ha utilizado la "full" por cuanto permite modelar la relación entre cada una de las variables.

5. Datos de entrenamiento, deben ser introducidos en un array de matrices

$$[D_1, D_2, D_3, ..., D_i, ..., D_n],$$

donde cada matriz D_i contiene los valores asociados a cada una de las observaciones de entrenamiento. Cada columna de la matriz D_i contiene los valores del frame respectivo.

A continuación, se enumeran las funciones utilizadas en esta tesis junto con los parámetros utilizados para la evaluación y entrenamiento de los HMM's. En el Anexo III, capítulo **??**, se puede obtener más información del uso de las mismas.

Para la inicialización se usó la función:
[pi, A, B, mu, Sigma] = init_mhmm(data, Q, M, cov_type)
Con:

- Q=4: Número de estados del modelo.
- M=5: Cantidad de mezclas Gaussianas en la distribución de probabilidad observación.
- cov_type='full': Tipo de matriz de covarianza, puede ser de tipo diagonal ('diag'), completa ('full') o esférica ('spherical').
- Data: Datos de entrenamiento. Correspondientes a los 124 individuos.

El algoritmo de inicialización devuelve:

- pi: Distribución inicial de probabilidades de cada estado.
- A: Matriz de distribución de probabilidades de transición.
- B: Matriz de distribución de probabilidades de observación.
- mu: Matriz con los valores medios asociados a las distribuciones Gaussianas.
- Sigma: Matriz con los desvíos estándar asociados a las distribuciones Gaussianas.

Con la función anterior ya se tienen las matrices iniciales para comenzar el entrenamiento. Los parámetros de entrada para el entrenamiento serán las salidas de la función anterior.

Para el entrenamiento se usó la función:
[ll_trace, pi, A, mu, sigma, B] = mhmm_em(data, pi0, A0, mu0, sigma0, B0,...)
Con:

- data: Datos de entrenamiento de los 124 individuos.
- pi0: Distribución inicial de probabilidades de cada estado, valor obtenido de la inicialización.
- A0: Matriz de distribución de probabilidades de transición, valor obtenido de la inicialización.
- B0: Matriz de distribución de probabilidades de observación, valor obtenido de la inicialización.
- mu0: Matriz con los valores medios asociados a las distribuciones Gaussianas, valor obtenido de la inicialización.
- Sigma0: Matriz con los desvíos estándar asociados a las distribuciones Gaussianas, valor obtenido de la inicialización.
- It=5 la llamada a la función es:

mhmm_em(data, pi0, A0, mu0, sigma0, B0, 'max_iter', It)

El algoritmo devuelve:

pi, A, mu, sigma, B: Matrices con los valores resultantes del entrenamiento del modelo.

Una vez terminada la fase de entrenamiento para los 124 individuos, se procede a realizar la fase de test. En ella se han tomado las matrices de salida del entrenamiento. Estas matrices se introducen como parámetros de entradas a la siguiente función, teniendo en cuenta que se les pasará los datos a testear que no fueron incluidos en la fase de entrenamiento.

Para la fase de test se utilizó:
[loglik, errors] = mhmm_logprob(data, pi, A, mu, sigma, B)

- data: Matriz con los datos a reconocer o testear. Corresponden a un individuo seleccionado al azar entre 1,..124.
- pi, A, mu, sigma, B: Matrices que representan el modelo HMM entrenado, las obtenidas anteriormente en la fase de entrenamiento.

El algoritmo devuelve:

- loglik: Verosimilitud logarítmica. Estima de la probabilidad de que la observación de entrada haya sido generada por el modelo.
- errors: Vector que indica en qué casos la verosimilitud logarítmica haya sido infinito.

Por tanto, en la variable loglik se tiene la probabilidad de que los datos usados para el test hayan sido generados por el modelo. Así, el resultado de los experimentos devuelve un conjunto de probabilidades.

Se realizan los experimentos en los que, el número de ejecuciones osciló entre 10 y 15 sobre los mismos datos para ver el comportamiento medio debido a su carácter estadístico. Las pruebas consisten en realizar una serie de entrenamientos de la persona en cuestión e ir pasándole después la secuencia de test para evaluar el método. En el caso de esta tesis son 124 probabilidades correspondientes a los 124 individuos, esto da una salida de probabilidades ordenadas de mayor a menor valor. Se escogen las cinco primeras probabilidades. Se escogen las cinco primeras probabilidades, ya que como se comentó, en la mayor parte de los casos la persona elegida quedaba con una probabilidad contenida entre estas cinco.

obtenidas correspondientes a cinco individuos y se va calculando la media de los valores obtenidos en los experimentos. El resultado es una probabilidad media de que la persona tomada para el test sea realmente la persona reconocida mediante el aprendizaje.

En la fig.4.10, se ha representado el porcentaje de aciertos por persona, teniendo en cuenta que el porcentaje se calcula sobre los valores de probabilidad que han obtenido las mejores posiciones entre las cinco primeras. En la primera de las gráficas se ha representado el porcentaje en el caso de persona normal, en la segunda en el caso de persona con bolso y en la tercera con abrigo. Se ha representado para todas y cada una de las personas existentes haciendo la media entre los resultados obtenidos para los diferentes ángulos y situación. Los porcentajes más altos se dan para el caso de la situación normal seguido de bolso y abrigo.

Los resultados de los experimentos muestran que para la situación normal del algoritmo HMM la media de los mejores valores alcanzados es para ángulos entre $72^\diamond$ - $108^\diamond$ (vista lateral) con el 97 %. Los peores resultados se obtienen para vistas frontales que fluctúan, aproximadamente, entre $0^\diamond$ - $18^\diamond$ y $144^\diamond$ - $180^\diamond$ con una media del 93 %. Como en el caso de los métodos estáticos, el mejor resultado se obtiene para el caso de una situación normal y vistas laterales.

En la fig. 4.11 se muestran los resultados de los experimentos para el caso de aplicar el método HMM. En el eje de abscisas, al igual que en el caso de los métodos estáticos, se situará el ángulo de la cámara y en el de ordenadas el porcentaje de

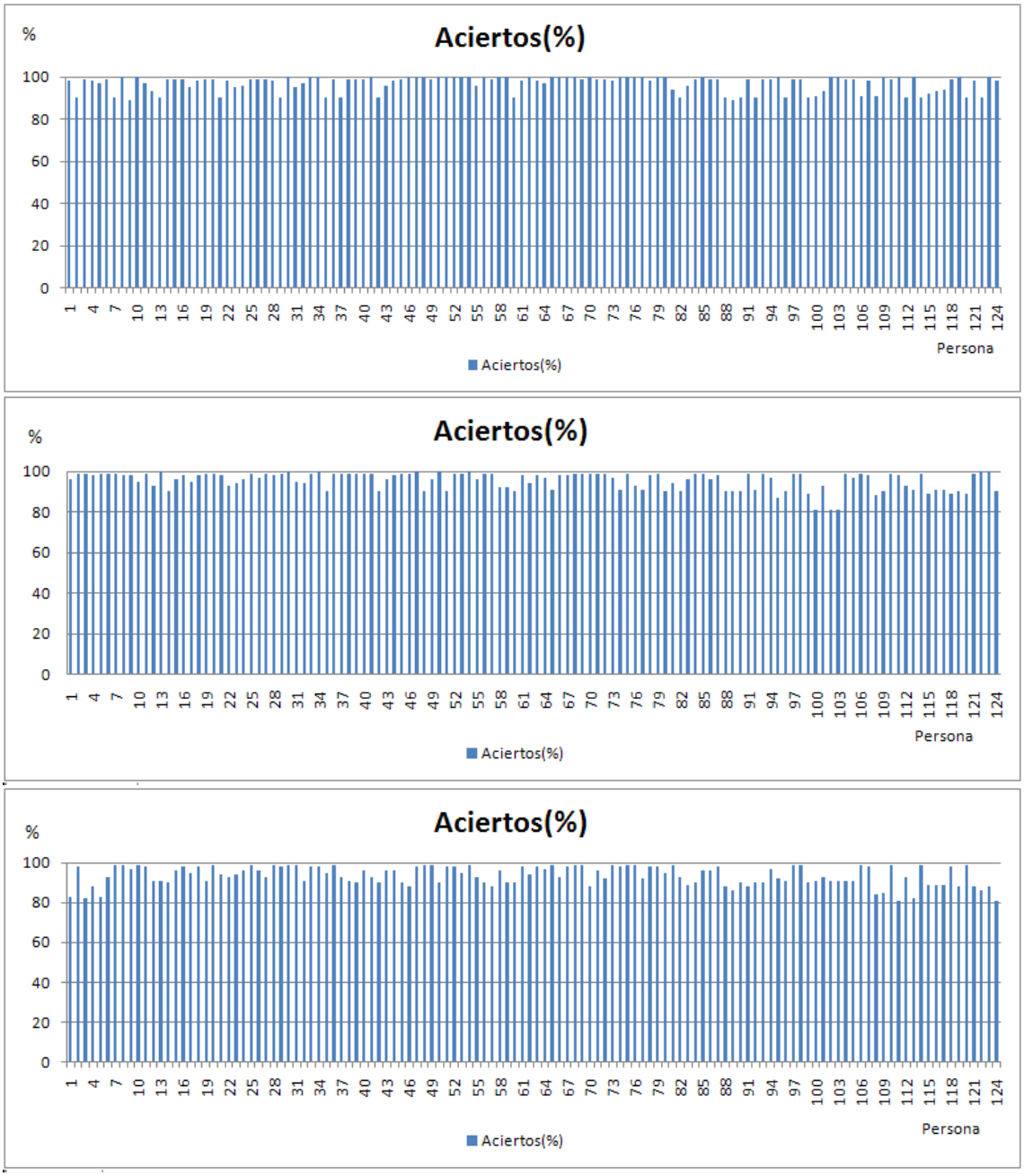


Figura 4.10: Resultados por persona: representación del número de aciertos por persona (Casos:NM, BG y CL), teniendo en cuenta que se considera acierto si la probabilidad de detectar al individuo correcto está entre los 5 primeros. En la primera de las gráficas se ha representado el porcentaje en el caso de persona normal, en la segunda en el caso de persona con bolso y en la tercera con abrigo.

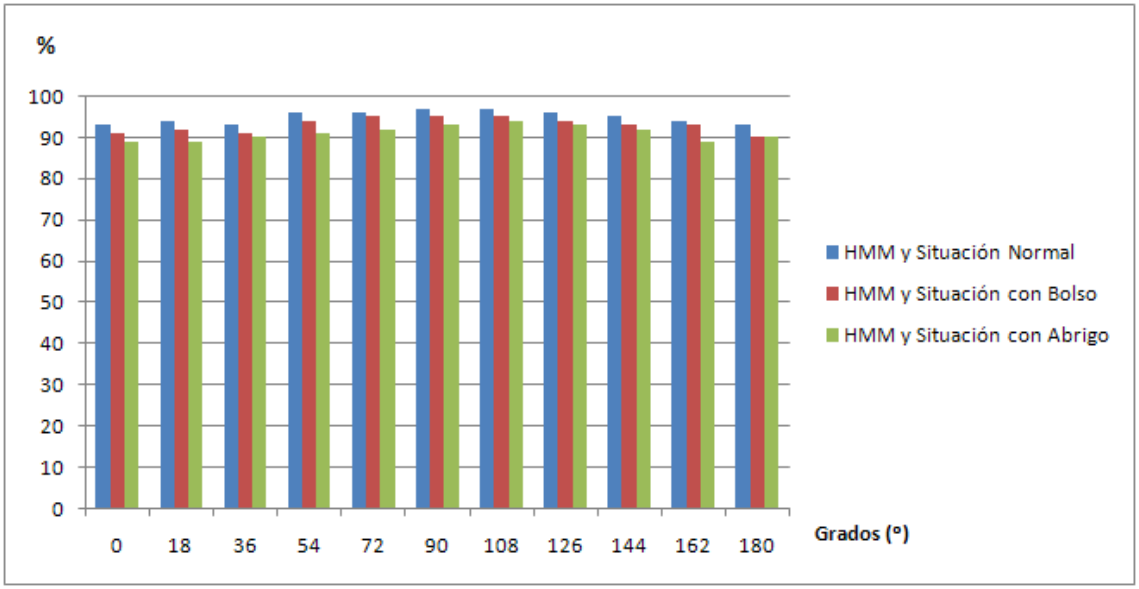


Figura 4.11: Resultados para el reconocimiento del Gait mediante el algoritmo HMM en diferentes situaciones y para diferentes ángulos de cámara. En el eje de abscisas, al igual que en el caso de los métodos estáticos, se situará el ángulo de la cámara y en el de ordenadas el porcentaje de aciertos, correspondiente a la persona a la persona analizada en cuestión, mediante diferentes ángulos. Por cada uno de los ángulos se representan las barras a diferentes colores con cada una de las situaciones en la que se encuentra (normal,bolso o abrigo).

aciertos correspondientes a dicho ángulo. Por cada uno de los ángulos se representan las barras a diferentes colores con cada una de las situaciones en la que se encuentra (normal,bolso o abrigo).

Se comprueba que, independientemente del método utilizado, los mejores resultados se obtienen para ángulos comprendidos entre 72^{o} a 108^{o}, es decir, vistas laterales. Por otro lado, los peores resultados se obtienen normalmente entre los rangos desde 0^{o} a 36^{o} y desde 144^{o} a 180^{o}, es decir, vistas frontales. Todo esto quiere decir, que los parámetros obtenidos del BB6-HM para el caso de vista lateral dan mucha más información, que para el caso frontal, lo cual es lógico si se comprueba que los parámetros que tienen que ver con la apertura de piernas (HC) o los aspavientos de los brazos, ayudan a la identificación del individuo.

También se observa, que en las secuencias con situación normal es donde se obtienen mejor resultados, seguidas de las de llevar bolso y por último las de abrigo. Al igual que el caso anterior, el problema puede residir en el hecho de que el abrigo elimina información de la silueta del individuo. Así, personas que se mueven con un abrigo ocultan parte del movimiento de las piernas. Lo mismo pasa con el hecho de llevar bolso, el movimiento de brazos queda limitado al ocultarse por el bolso.

Algoritmo	Situación	Max Valores (grados)	% resultados (media)	Min valores (grados)	% resultados (media)
HMM	N	72°, 90°, 108°	97%	0°,36° 144°,180°	93%
	B	72°, 90°, 108°	95%	18°,162°	91%
	A	108°	93%	0°,18°, 36°,162°,180°	89%

Tabla 4.7: Tabla que recoge los resultados de los experimentos según el ángulo de la cámara para los HMM's. La tabla se agrupa según el tipo de situación de la imagen, N: normal, B: bolso, A: abrigo. La columna "Max Valores", corresponde ángulos de mejores resultados (grados) y " % resultados" será los valores obtenidos (%), para esos ángulos. De igual modo, la columna "Min Valores" corresponde los ángulos de peores resultados (grados) y " % resultados" será el valor obtenido (%) para esos ángulos.

A continuación se muestra una tabla resumen (Tabla 4.7), al igual que en el caso de los métodos estáticos, de todas las gráficas anteriores. La tabla se agrupa según el tipo de situación de la imagen, N: normal, B: bolso, A: abrigo. La columna "Max Valores", corresponde los valores máximos del ángulo de la cámara (grados) y " % resultados" será los valores obtenidos (en porcentaje), para esos ángulos. De igual modo, la columna "Min Valores" corresponde los valores mínimos del ángulo de la cámara (grados) y " % resultados" será el valor obtenido (en porcentaje) para esos ángulos.

4.5. Resumen de los experimentos realizados

A la vista de los resultados del apartado anterior, se deduce que los mejores resultados se han obtenido para métodos de Máquinas de Vectores de Soporte (SVM) en el caso de métodos estáticos, sin embargo, los métodos dinámicos, Modelos Ocultos de Markov (HMM), son los que obtienen en general los mejores resultados de todos los métodos empleados en la Tesis.

A continuación, se muestra la Tabla 4.8, comparativa entre los modelos estáticos (árboles decisión, SVM, etc) y los modelos dinámicos (HMM). En ella se representan los datos medios por tipo de situación (normal, bolso, abrigo), para el caso más favorable y para el menos.

Se aprecia que los métodos dinámicos proporcionan en todas las situaciones mejores resultados que los estáticos, tanto para las medias de todos los mejores resultados obtenidos como la de los peores.

En la Tabla 4.9, se muestra la media de los resultados obtenidos, según modelos estáticos y dinámicos, por ángulos. En ella se observa que los ángulos comprendidos

Media de los mejores resultados	Situación	Estáticos	HMM
	Normal	95,25 %	96,86 %
	Bolso	93,15 %	95,71 %
	Abrigo	91,2 %	93,67 %
Media de los peores resultados			
	Normal	86,55 %	93,5 %
	Bolso	84 %	92,3 %
	Abrigo	82,5 %	91,13 %

Tabla 4.8: Comparativa entre los modelos estáticos (árboles decisión, SVM , etc) y los modelos dinámicos (HMM), respecto a la media de los resultados obtenidos y situación dada.

Ángulo	Estáticos (%)	Dinámicos (%)
0º	84,91	90
18º	84,66	91
36º	84,41	91,33
54º	91,25	92
72º	92,66	94,33
90º	94,58	96
108º	93,33	95,33
126º	91,33	94,66
144º	86	93
162º	84,5	91
180º	83,91	89

Tabla 4.9: Media de los resultados obtenidos en porcentaje, según modelos estáticos y dinámicos, por ángulos.

entre 90º-124º, dan mejores resultados, tal y como, ya se había comentado en las gráficas anteriores. A medida que se va alejando de estos ángulos se va descendiendo gradualmente. Encontrándose los valores más bajos en los extremos (0º-180º).

En la Tabla 4.10se muestra un resumen con la media del porcentaje con los resultados obtenidos por los diferentes métodos para la evaluación los parámetros definidos en la subsección 4.3.1. En el caso de "vista_lateral_frontal" se evalúa si el individuo se mueve de modo lateral o frontal y en la "vista" se representan los ángulos de la cámara (0º hasta 180º). Se aprecia, que para las pruebas que se han realizado, los mejores resultados se han obtenido para el método SVM.

También se ha realizado un estudio de los tiempos, tanto de los tiempos de ejecución que tarda al procesar las imágenes y calcular los parámetros del modelo, como los tiempos de entrenamiento y test. Para el cálculo de los tiempos de ejecución,

Parámetro\Método	J48	META	MLP	SVM
Vista	86,6%	90,8%	88,6%	92,85%
Vista_Lateral_Frontal	99,2%	99,26%	99,1%	99,5%

Tabla 4.10: Tabla con los resultados de la evaluación de los casos 1,2 y 4. Los mejores resultados se han obtenido para las SVM.

mediante reglas heurísticas, al evaluar los parámetros del modelo se ha utilizado la sentencia "clock", la cual devuelve el número de segundos que ha dedicado el procesador a la ejecución de la sentencia considerada. Así, se ha obtenido un tiempo de procesado de la imagen en el que se ha considerado el tiempo de umbralización (para filtrar los pixeles de una imagen teniendo en cuenta para su discriminación la intensidad de cada pixel), el tiempo de búsqueda (para encontrar el contorno a una imagen segmentada) y el tiempo de dibujado del contorno y el tiempo para eliminar la distorsión de la imagen, entre otros . También se ha calculado el tiempo de apertura de la imagen, así como el tiempo de escritura de los ficheros con los paramétros obtenidos. Con todo esto se ha obtenido un conjunto de resultados de los valores medios de todas las muestras analizadas. En la Tabla 4.11 se muestran los tiempos medios (en segundos) de ejecución en el cálculo del procesado de imagen, los parámetros del modelo y los tiempos de apertura de imagen y escritura de ficheros. También se muestra el tiempo total de cómputo sumando todas las columnas anteriores.

Como ya se explicó en las subsecciones 4.4.1 y 4.4.2, se realizaron experimentos con un conjunto de atributos que inicialmente fueron de 119 y los tiempos de cómputo para el entrenamiento resultaban muy elevados. Por esta razón se aplicó PCA's con el fin de ayudar en la selección de las variables del vector de características. Las PCA's transforman un conjunto de variables (atributos) correladas en un conjunto menor de variables no correladas. Estas se aplicaron sobre el conjunto de datos obteniéndose vectores que fuesen capaces de explicar el 95 % de la variación de los datos. Además de utilizar las PCA's, se utilizarón otros métodos como los árboles de decisión ya que muestran de forma gráfica desde las variables más significativas a las menos en forma de árbol. Todo esto sirvió para reducir las variables, llegándose a 18 para el reconocimiento del Gait mediante métodos estáticos o 18 para el caso de los dinámicos. y obteniéndose mejores tiempos en todos y cada uno de los clasificadores utilizados en esta tesis. La Tabla 4.12 muestra los tiempos medios de entrenamiento

Tiempo procesado imagen y búsqueda de contorno	Tiempo medio en el cálculo de cada parámetro del modelo	Tiempo apertura imagen y escritura ficheros	Tiempo Total
0.00446	0.000078	0.033252	0.03779
0.00432	0.000047	0.022789	0.02715
0.00435	0.000063	0.019649	0.02406
0.00448	0.000078	0.021663	0.02622
0.00409	0.000048	0.029268	0.03340
0.00434	**0.0000628**	**0.025324**	**0.02972**

Tabla 4.11: Tiempos medios de ejecución (segundos) en el cálculo del procesado de imagen, los parámetros del modelo y los tiempos de apertura de imagen y escritura de ficheros. También se muestra el tiempo total de cómputo sumando todas las columnas anteriores. La última fila de la tabla muestra la media por columna de todos los valores mostrados en la tabla.

Método	Tiempo Medio de Entrenamiento	Tiempo Medio de Test
J48	5233.76	10.7
META	13561.93	21.39
MLP	19510	26.91
SVM	20179	23.32
HMM	14503	5.82

Tabla 4.12: Tiempos medios de entrenamiento y de test (segundos) por cada uno de los métodos utilizados en esta tesis.

y de test (en segundos) por cada uno de los metodos utilizados en esta tesis. El método META, corresponde a la media de los resultados obtenidos entre todos los metaclasificadores (boosting, bagging y stacking), ya que dieron tiempos muy altos en conjunto, que superaron al resto, y por eso se agruparon dentro de la misma etiqueta. Los resultados muestran que que las SVM y MLP dieron peores resultados en cuanto a tiempos de entrenamiento y test , siendo en conjunto muy similares. Los HMM contituyen un caso aparte en cuanto a la comparación de tiempos, puesto que ni la herramienta utilizada no es la misma ni el tipo de método utilizado (dinámico).

Capítulo 5

Aportaciones y trabajo futuro

La aportación de esta tesis es la presentación de un nuevo modelo (BB6-HM), orientado principalmente a vigilancia, el cual ofrece un número de ventajas comparado con otros. Este modelo, es capaz de monitorizar, en tiempo real y con una carga computacional mínima, una importante cantidad de estos eventos primitivos relacionados con humanos, además, es modular e invariante a la escala y al punto de vista, lo que lo hace muy útil, para la descripción de actividades más complejas en multitud de tareas de vigilancia

La modularidad del sistema facilita la definición de nuevos estados primitivos y eventos. Por tanto, el sistema es fácilmente extensible, de modo que es posible integrar nueva información para obtener una descripción más rica del humano y así incrementar tanto la confianza como el número de eventos reconocidos.

El hecho de que el sistema esté orientado desde el inicio a vigilancia justifica la amplia cantidad de eventos que se pueden extraer del mismo para este fin.

Los parámetros del BB6-HM permiten determinar, con los diferentes métodos de aprendizaje, los parámetros tanto de las vistas frontales como laterales. Si bien se ha visto que los parámetros obtenidos para cada caso son diferentes y los resultados también.

Por tanto, como dos características claves, se puede decir que esta tesis aporta un modelo intuitivo y robusto. Intuitivo porque se ejemplifica con el uso de análisis simples de la variación de algunos de los parámetros del modelo para reconocer eventos y situaciones simples. La selección de los parámetros y sus valores se realiza por métodos heurísticos. Robusto porque se ejemplifica con el uso de algoritmos de aprendizaje supervisado a partir de datos para la selección del subconjunto de parámetros y su configuración para reconocer eventos y situaciones más complejas

En futuros trabajos, este modelo se podrá aplicar a la detección de nuevos eventos primitivos y atributos visuales e integrarlo en un sistema automático que permita

monitorizar actividades humanas (control de acceso, control de presencia, monitorización de actividades, detección de situaciones de alarma, etc).

Además, la información proporcionada por el BB6-HM desde puntos de vista diferentes podría ser integrada a fin de obtener una descripción 3D del humano y así aumentar la fiabilidad del sistema y reconocer un mayor número de acontecimientos.

Otra línea de trabajo utilizando este modelo sería la detección de malas segmentaciones en la imagen con el fin de mejorar la misma.

Capítulo 6

Conclusiones

En este trabajo se ha presentado el modelo BB6-HM, un nuevo modelo de representación de la dinámica de los humanos orientado a la caracterización de humanos para la monitorización de sus actividades en tareas de vigilancia . Este modelo fácil ... que es capaz de monitorizar, en tiempo real y con una carga computacional mínima, una importante cantidad de estos eventos primitivos relacionados con humanos, además, es modular e invariante a la escala y al punto de vista, lo que lo hace muy útil, para la descripción de actividades más complejas en multitud de tareas de vigilancia. La modularidad del sistema facilita la definición de nuevos estados primitivos y eventos. Por tanto, el sistema es fácilmente extensible, de modo que es posible integrar nueva información para obtener una descripción más rica del humano y así incrementar tanto la confianza como el número de eventos reconocidos. Es un modelo potente, con multitud de parámetros que pueden ser utilizados para la caracterización y reconocimiento de gran cantidad de eventos y situaciones que puedan ser descritos o identificados a partir de la silueta del humano.

En esta tesis se han explorado distintos tipos de información que podemos extraer del análisis:

- Información sobre la localización de las partes del cuerpo.
- Información sobre el movimiento del humano: velocidad, dirección respecto a la cámara y periodicidad del movimiento.
- Reconocimiento de situaciones primitiva*s: postura y especiales, como oclusiones y acarreo de objetos.*
- Reconocimiento de la persona por la manera de andar ("Gait").

Ha quedado patente la cantidad de información orientada a la tarea de vigilancia o monitorización de actividades que se puede extraer del modelo BB6-HM (oclusiones,

acarrear objetos, dejarlos, cogerlos, agacharse, levantarse, ...etc), cada una a partir de un subconjunto pequeño de parámetros del modelo y, por tanto, con un coste computacional bastante pequeño y evidencia de que puede ser utilizado en tiempo real.

El modelo está limitado al reconocimiento de eventos y situaciones realizados por una sola persona, pero la filosofía del modelo, esto es, analizar la división del blob en bloques del blob y su caracterización visual por medio de un conjunto de parámetros que describen sus propiedades de color y características espacio-temporales, es fácilmente aplicable a la interacción de grupos de personas o de otro tipo de objetos (vehículos, perros, etc.)

Por otro lado, se ha utilizado el modelo el cual basándose en el conjunto de la postura y el movimiento al caminar (gait), ha sido capaz de reconocer personas dentro de un conjunto dado, obteniéndose muy buenos resultados con un modelo fácil de tratar y modelar, y, con el mismo pequeño coste computacional. Sin embargo, este es un problema aun sin resolver definitivamente, pues los resultados obtenidos son muy dependientes de la información proporcionada por la silueta del humano, la cual puede verse afectada por la segmentación previa y por todo aquello que modifique la silueta del humano (vestimenta, equipaje, etc.).

Los resultados de los experimentos han demostrado la capacidad del modelo para la detección de eventos primitivos y visuales relacionados con la vigilancia. Estos eventos y situaciones de interés pueden ser descritos mediante modelos estáticos y/o dinámicos, mediante un análisis cualitativo y cuantitativo de los parámetros del modelo, definiendo heurísticas, o mediante aprendizaje supervisado a partir de datos. En el caso de utilizar aprendizaje es muy importante un buen diseño del experimento para identificar las características relevantes del modelo y un buen preprocesado inicial que elimine información incorrecta que puede confundir a los algoritmos de aprendizaje. Los resultados demuestran que la información se puede extraer de un número pequeño de parámetros. A la vista de los resultados, los mejores resultados se han obtenido con el método de las Máquinas de Vectores de Soporte (SVM) para el entrenamiento de modelos estáticos. En el caso de entrenamiento de modelos dinámicos, sólo se ha probado con modelos ocultos de markov. La conclusión general es que modelos estáticos son suficientes para modelar situaciones y eventos de corta duración, donde la variación temporal no es significativa, pero que, a medida que la duración de la situación se alarga, los modelos dinámicos obtienen mejores resultados.

Logros del modelo:

- Monitorización, en tiempo real.

- Carga computacional mínima.
- Detección de cantidad de estos eventos primitivos relacionados con humanos.
- Modular e invariante a la escala y al punto de vista.
- El sistema es fácilmente extensible.
- Proporciona multitud de parámetros que pueden ser utilizados para la caracterización y reconocimiento de gran cantidad de eventos y situaciones que puedan ser descritos o identificados a partir de la silueta del humano.

Trabajos futuros:

- Detección de nuevos eventos primitivos y atributos visuales e integrarlo en un sistema automático que permita monitorizar actividades humanas (control de acceso, control de presencia, monitorización de actividades, detección de situaciones de alarma, etc).

- La información proporcionada por el BB6-HM desde puntos de vista diferentes podría ser integrada a fin de obtener una descripción 3D del humano y así aumentar la fiabilidad del sistema y reconocer un mayor número de acontecimientos.
- Detección de malas segmentaciones en la imagen con el fin de mejorar la misma.

Bibliografía

Allen, B., Curless, B., & Popovic, Z. (2002). Articulated Body Deformation from Range Scan Data. *Proc. ACM SIGGRAPH 21(3)*, (pp. 612–619).

Andersen, B., Dahl, T., Iversen, M., Pedersen, M., & Søndergaard, T. (2001). Complex Human Motion Capture. *Computer Vision and Image Understing 81(3)*, (pp. 231–268).

Arici, T. (2012). Introduction to Programming with Kinect: Understanding Hand/arm/head Motion and Spoken Commands. *Signal Processing and Communications Applications Conference (SIU).*

Aswatha, K., Chatterji, M., Mukherjee, B. N., & Das, J. (1996). Representation of 2D and 3D Binary Images Using Medial Circles and Spheres. *Int. J. Pattern Recognition. 10 (4)*, (pp. 365–387).

Azarbayejani, A., Wren, C. R., & Pentland, A. P. (1997). Pfinder:Real-Time Tracking of the Human Body. *IEEE Trans. Pattern Anal. Mach. Intell. 19(7)*, (pp. 780–785).

Badler, N., O'Rourke, J., & Toltzis, H. (1979). A Spherical Representation of a Human Body for Visualizing Movement. In *IEEE Proceedings 67(10)* (pp. 1397–1403).

Bao, L. & Intille, S. S. (2004). Activity Recognition from User-annotated Acceleration Data. *Pervasive Computing vol. LNCS 3001*, (pp. 1–17).

Basu, S., Essa, I., & Pentland, A. (1996). Motion Regularization for Model-Based Head Tracking. *Perceptual ComputingSection, Media Laboratory Massachusetts Institute of Technology Cambridge MA. U.S.A*, (pp. 611–616).

Baum, L. E., Petrie, T., Soules, G., & Weiss, N. (1970). A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains. *The Annals of Mathematical Statistics*, 41(1), 164–171.

Begler, C. & Malik, J. (1997). Video Motion Capture. *Computer Science Division and Univ. Of California Berkeley Berkeley CA 9*, (pp. 4720–1776).

Bobick, A. (1997). Movement, Activity and Action: The Role of Knowledge in the Perception of Motion. *Philosophical Trans. Royal Soc. London, 352.*, (pp. 1257–1265).

Bobick, A. & Davis, J. (2001). The Recognition of Human Movement Using Temporal Templates. *IEEE Trans on Pattern Analysis and Machine Intelligence vol.23 , no. 3*, (pp. 257–267).

Bobick, A. & Wilson, A. (1995). A State-based Technique for the Summarization and Recognition of Gesture. *Proc. of Intl. Conf. on Computer Vision. Cambridge*, (pp. 382–388).

Bouchrika, I. & Nixon, I. M. (2007). Model-based Feature Extraction for Gait Analysis and Recognition. In *Model-based Imaging and Rendering Image Analysis Graphical special Effects* (pp. 150–160).

Bouchrika, I. & Nixon, M. (2008). Exploratory Factor Analysis of Gait Recognition. *In and 8th IEEE International Conference on Automatic Face Gesture Recognition Amsterdam The Netherls.*

Boulay, B., Bremond, F., & Thonnat, M. (2003). Human Posture Recognition in Video Sequence. *VS-PETS (Visual Surveillance and Performance Evaluation of Tracking Surveillance)*, (pp. 23–29).

Boulay, B., Bremond, F., & Thonnat, M. (2005). Posture Recognition with a 3D Human Model. *ICDP (Imaging for Crime Detection and Prevention)*, (pp. 135–138).

Boulgouris, N. & Chi:, Z. (2007). Human Gait Recognition Based on Matching of Body Components. *Pattern Recognition, 40*, (pp. 1763–1770).

Boyd, J. E. & Little, J. (2005). Biometric Gait Recognition. *LNCS 3161 , Springer*, (pp. 19–42).

Bradski, G. (1998). Real time face and object tracking as a component of a perceptual user interface. In *Applications of Computer Vision, 1998. WACV '98. Proceedings., Fourth IEEE Workshop on* (pp. 214–219).

Brand, M., Oliver, N., & Pentland, A. (1997). Coupled hidden Markov models for complex action recognition. In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on* (pp. 994–999).

Branko, R., Arulampalam, S., & Gordon, N. (2004). Beyond the Kalman Filter: Particle Filters for Tracking Applications. *Artech House Radar Library.*

Bregler, C. (1997). Learning and Recognition Human Dynamics in Video Sequrence. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'97)*, (pp. 568–574).

Bregler, C. & Malik, J. (1998). Tracking People with Twists and Exponential Maps. *IEEE International Conference on Computer Vision and Pattern Recognition, Santa Barbara, CA 8.*

Broggi, A., Bertozzi, M., Fascioli, A., & Sechi, M. (2000). Shape-based Pedestrian Detection. In *Procs. IEEE Intelligent Vehicles Symposium.* (pp. 215–220).

Bruzzone, L., Cossu, R., & Vernazza, G. (2004). Detection of Land-cover Transitions by Combining Multidate Classifiers. *Pattern Recognition. Lett, 25*, (pp. 1491–1500).

Calinon, S., Guenter, F., & Billard, A. (2005). Goal-directed Imitation in a Humanoid Robot. *Proc IEEE Int Conf. Robotics and Automation ICRA05 Barcelona Spain*, (pp. 299–304).

Cascia, M. L., Isidoro, J., & Sclaroff, S. (1998). Head Tracking Via Robust Registration in Texture Map Images. In *Proc. CVPR 98", June . 508.*

CASIA, Chinese Academy of Sciences (2005). CASIA Gait Database. http://www.cbsr.ia.ac.cn/english/Gait Databases.asp.

Chang, I. C. & Huang, C. L. (2000). The Model-based Human Body Motion Analysis System. *Image and Vision Computing 18(14)*, (pp. 1067–1083).

Chen, C., Liang, J., Zhao, H., Hu, H., & Tian, J. (2009). Factorial HMM and Parallel HMM for Gait Recognition. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 39(1), 114–123.

Chen, Z. & Lee, H. (1992). Knowlegde-guided Visual Perception of 3D Human Gait from a Single Image Sequence. *IEEE transactions on Systems, Man Cybernetics 22 (2)*, (pp. 336–342).

Cheng, F., Christmas, W. J., & Kittler, J. (2002). Recognising Human Running Behaviour in Sports Video Sequences. *International Conference on Pattern Recognition, Quebec Canada.*, (pp. 11–15).

Cheng, L., Chang, I. C., & Huang, C. L. (1996). Ribbon-based Motion Analysis of Human Body Movements. *Proc. of Intl. Conf. on Pattern Recognition. Vienna*, (pp. 436–440).

Cheng, L., Chang, I. C., Huang, C. L., & Huang, W. L. (1998). The Body Signature-based Motion Analysis for Walking Human. *Conference on Computer Vision and Graphics Image Processing Taiwan.*

Chenyang, X. & Prince, J. (1998). Snakes, Shapes, and Gradient Vector Flow. *Image Processing and IEEE Transactions.*, (pp. 359–369).

Cheung, K., Kanade, T., Bouguet, J. Y., & Holler, M. (2000). A Real Time System for Robust 3D Voxel Reconstruction of Human Motions. In *Proc. of Computer Vision and Pattern Recognition vol. 2* (pp. 2714–2720).

Cho, C., Chao, W., Lin, W. H., & Y.Chen (2009). A Vision-based Analysis System for Gait Recognition in Patients with Parkinson's Disease. *Expert Systems with Applications, 36(3)*, (pp. 7033–7039).

Cohen, I. (2004). 3D Hand and Fingers Reconstruction. *The integrated Media System Center (IMSC) at the University of Southen California (USC).*

Cohen, I. & Lee, M. W. (2002). 3D Body Reconstruction for Immersive Interaction. *Second InternationalWorkshop on Articulated Motion and Deformable Objects, Palma de Mallorca, Spain*, (pp. 21–23).

Cohen, I. & Li, H. (2003). Inference of Human Postures by Classification of 3D Human Body Shape. *IEEE International Workshop on Analysis and Modeling of Faces Gestures*, (pp. 74–81).

Cohen, I. & Medioni (1999). Detecting and Tracking Moving Objectsfor Video Surveillance. In *Computer Vision and Pattern Recognition, Fort Collins Colorado June.*

Collins, Lipton, Kanade, Fujiyoshi, Duggins, Tsin, Tolliver, Enomoto, & Hasegawa (2000). *A System for Video Surveillance and Monitoring: VSAM Final Report.* Technical report. CMU-RI-TR-Robotics Institute, Carnegie Mellon University May.

Cortes, C. & Vapnik, V. (1995). Support-vector Networks. *Machine Learning , vol. 20*, (pp. 273–297).

Csaba, B., Fruhstuck, B., Bischof, H., & Kropatsch, W. (2005). Model-Based Occlusion Handling for Tracking in Crowded Scenes. *Joint Hungarian-Austrian Conference on Image Processing and Pattern Recognition 5th KÉPAF 29th AGM Workshop*, (pp. 227–234).

Darrell, T., Maes, P., Blumberg, B., & Pentland, A. P. (1994). A Novel Environment ForSituated Vision and Behavior. *In Workshop for Visual on Behaviors at CVPR-94*, (pp. 68–72).

Davis, J. & Gao, H. (2003). Recognizing Human Action Efforts: An Adaptive Three-mode Pca Framework. *Proc. Int. Conf. On Computer Vision (ICCV03). Vol 2.*, (pp. 1463–1469).

Davis, L., Philomin, V., & Duraiswami, R. (2000). Tracking Humans from a Moving Platform. *Computer Vision Laboratory Institute for Advanced Computer StudiesUniversity of Maryland College Park MD 20742 , USA vol 4*, (pp. 171–178).

Davis, R. I. A. & Lovell, B. C. (2003). Comparing and evaluating HMM ensemble training algorithms using train and test and condition number criteria. *Pattern Anal. Appl.*, 6(4), 327–336.

Delaitre, V., Laptev, I., & Sivic, J. (2010). Recognizing Human Actions in Still Images: A Study of Bag-of-features and Part-based Representations. In *Proceedings of the British Machine Vision Conference*.

Deng, L. & Erler, K. (1991). Microstructural Speech Units and Their HMM Representations for Discrete Utterance Speech Recognition. *in ICASSP-91 , 193-196.*

Deng, L. & Sun, D. (1994). Phonetic Classification and Recognition Using HMM Representation of Overlapping Articulator Features for All Classes of English Sounds. *in ICASSP-94 , 45-47.*

Deutscher, J., Blake, & I.Reid (2000). Articulated Body Motion Capture by Annealed Particle Filtering. In *Computer Vision and Pattern Recognition Proceedings. IEE Conference on Vol: 2* (pp. 126–133).

Di Bernardo, E., Goncalves, L., & Perona, P. (1996). Monocular tracking of the human arm in 3D: real-time implementation and experiments. In *Pattern Recognition, 1996., Proceedings of the 13th International Conference on*, volume 3 (pp. 622–626).

Do, Y. (2005). Tracking People in Video Camera Images Using Neural Networks. In *ICIC'05 Proceedings of the 2005 international conference on Advances in Intelligent Computing - Volume Part I.* (pp. 301–309).

Efros, A., Berg, A. C., Mori, G., & Mali, J. (2003). Recognizing Action At a Distance. *Proc. Int. Conf. on Computer Vision, (ICCV03). Vol 2.*, (pp. 726–734).

Elgammal, A., Shet, V., Yacoob, Y., & Davis, L. (2003). Learning Dynamics for Exemplar Based Gesture Recognition. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition Madinson WI USA. Vol 1.*, (pp. 571–578).

Eng, H. L., Toh, K. A., Kam, A. H., Wang, J., & Yau, W. Y. (2003). An Automatic Drowning Detection Surveillance System for Challenging Outdoor Pool Environments. *Proc. Int. Conf. on Computer Vision (ICCV03). Vol 1*, (pp. 532–539).

Ermes, M., Parkka, J., & Cluitmans, L. (2008). Advancing from Offline to Online Activity Recognition with Wearable Sensors. *Engineering in Medicine and Biology Society. 30th Annual International Conference of the IEEE*, (pp. 4451–4454).

Essa, I., Basu, S., Darrell, T., & Pentland, A. (1996). Modeling, Tracking and Interactive Animation of Faces and Head Using Input from Video. In *In: Proceedings of Computer Animation, Geneva June* (pp. 68–79).

Fanti, C., Zelnik-Manor, L., & Perona, P. (2005). Hybrid Models for Human Motion Recognition. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition CVPR05. Vol 1.*, (pp. 1166–1173).

Fathi, A. & Mori, G. (2008). Action Recognition by Learning Mid-level Motion Features. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR'08), Anchorage AK* (pp. 1–8).

Fine, S., Singer, Y., & Tishby, N. (1998). The Hierarchical Hidden Markov Model: Analysis and Applications. *Machine Learning*, 32(1), 41–62.

Folgado, E., Rincón, M., Bachiller, M., & Carmona, E. J. (2009). A Block-Based Human Model for Visual Surveillance. In *Proceedings of the 3rd International Work-Conference on The Interplay Between Natural and Artificial Computation: Part II: Bioinspired Applications in Artificial and Natural Computation*, IWINAC '09 (pp. 208–215). Berlin, Heidelberg: Springer-Verlag.

Folgado, E., Rincón, M., Carmona, E. J., & Bachiller, M. (2011). A block-based model for monitoring of human activity. *Neurocomput.*, 74(8), 1283–1289.

Foresti, G., Micheloni, C., Snidaro, L., & Marchiol, C. (2003). Face Detection for Visual Surveillance. In *Image Analysis and Processing 2003. Proceedings. 12th International Conference onSept. 115-120* (pp. 7–19).

F.Porikli & Tuzel, O. (2003). Human Body Tracking by Adaptive Background Models and Mean-Shift Analysis. *Mitsubishi Electric Research Labs and Murray Hill NJ 07974 , USA*.

Fuentes, L. M. & Velastin, S. A. (2006). People Tracking in Surveillance Applications. *In Image and Vision Computing 24(11) Elsevier November*, (pp. 1165–1171).

Fujiyoshi, H. & Lipton, A. J. (1998). Real-time human motion analysis by image skeletonization. *Proc. of IEEE Workshop on Applications of Computer Vision.*, (pp. 15–21).

Fujiyoshi, H. & Lipton, A. J. (2004). Real-Time Human Motion Analysis by Image Skeletonization. *IEICE (Engineering Sciences Society Communications Society Electronics Society Information) Transactions*, (pp. 113–120).

Fung, G. & Stoeckel, J. (2007). SVM Feature Selection for Classification of SPECT Images of Alzheimer's Disease Using Spatial Information. *Knowledge and Information Systems. 11(2)*, (pp. 243–258).

Gallant, S. (1990). Perceptron-based Learning Algorithms. *IEEE Transactions on Neural Networks, vol. 1 , no. 2*, (pp. 179–191).

Gao, J., Hauptmann, A. G., & Wactlar, H. D. (2004). Combining Motion Segmentation with Tracking for Activity Analysis. *International Conference on Automatic Face and Gesture Recognition Seoul Korea.*, (pp. 699–704).

García-Rojas, A., M.Gutiérrez, & Thalmann, D. (2008). Visual Creation of Inhabited 3D Environments. *The Visual Computer 24, 719-726*, (pp. 7–9).

Gavrila, D. (1999). The Analysis of Human Motion and Its Application for Visual Surveillance. *Fort Collins, U.S.A*, (pp. 3–5).

Gavrila, D. & Davis, L. (1996). 3D Model-based Tracking of Human in Action a Multiview Approach. *Proc Conf. Computer Vision and Pattern Recognition*, (pp. 73–80).

Gavrila, D. & Philomin, V. (1999). Real-Time Object Detection for Smart Vehicles. *DaimlerChrysler Research, vol. 1*, (pp. 87–93).

GBharatkumar, A., Daigle, K., Pandy, M., Cai, Q., & Aggarwal, J. (1994). Lower Limb Kinematics of HumanWalking with the Medial Axis Transformation. In *Workshop on Motion of Non-Rigid and Articulated Objects Austin Texas USA*.

Ghahramani, Z. & Jordan, M. I. (1997). Factorial Hidden Markov Models. *Mach. Learn.*, 29(2-3), 245–273.

Girondel, V., Bonnaud, L., & Caplier, A. (2006). A Human Body Analysis System. *EURASIP (European Association for Signal Processing), Journal on Applied Signal Processing.*, (pp. 1–18).

Goffredo, M., Seely, R., Carter, J., & Nixon, M. (2008). Markerless View Independent Gait Analysis with Self-camera Calibration. *IEEE International Conference on Automatic Face and Gesture Recognition. Amsterdam The Netherls*, (pp. 1–6).

Gonzalez, J., Varona, J., Roca, F. X., & Villanueva, J. J. (2002). ASpaces: Action Spaces for Recognition and Synthesis of Human Actions. *AMDO*, (pp. 189–200).

Green, R. & Guan, L. (2004). Continuous human activity recognition. In *Control, Automation, Robotics and Vision Conference, 2004. ICARCV 2004 8th*, volume 1 (pp. 706–711).

Group, M. L. (2012). Weka 3: Data Mining Software in Java. http://www.cs.waikato.ac.nz/ml/weka/.

Guo, B. & Nixon, M. (2009). Gait Feature Subset Selection by Mutual Information. *IEEE Trans. Systems and Man Cybernetics 39(1)*, (pp. 36–46).

Gutta, S., Huang, J., V.Kakkad, & Wechsler, H. (1998). Face Surveillance. *Department of Computer Science George Mason University. ICCV*, (pp. 646–651).

Hancock, P., Bruce, V., & Burton, A. (1998). A Comparison of Two Computer-based Face Recognition Systems with Human Perceptions of Faces. *Vision Research , vol 38*, (pp. 2277–2288).

Haritaoglu, I., Harwood, D., & Davis, L. (2000). W4: real-time surveillance of people and their activities. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8), 809–830.

Haritaoglu, I., Harwood, D., & Davis, L. S. (1998a). Ghost: A Human Body Uter Part Labeling System Using Silhouettes. *International Conference on Pattern Recognition vol 1*, (pp. 77–82).

Haritaoglu, I., Harwood, D., & Davis, L. S. (1998b). W 4: Who? When? Es- Where? What? A Real Time System for Detecting and Tracking - People. *International Conferenceon Automatic Face- and Gesture- yda Recognition Nara Japón.*

Harwood, I. & Davis, D. (1998). System for Detecting and Tracking People. *3rd Face I& Gesture Recog. Conf.*

He, Z. & Jin, L. (2008). Activity Recognition from Acceleration Data Using an Model Representation and Svm. *International Conference on Machine Learning and Cybernetics vol. 4*, (pp. 2245–2250).

He, Z. & Jin, L. (2009). Activity Recognition from Acceleration Data Based on Discrete Consine Transform and Svm. *IEEE International Conference on Systems and Man Cybernetics*, (pp. 5041–5044).

He, Z., Liu, Z., Jin, L., Zhen, L.-X., & J.-C., H. (2008). Weightlessness Feature: a novel feature for single tri-axial accelerometer based activity recognition. *19th International Conference on Pattern Recognition*, (pp. 1–4).

Heisele, B. & Wöhler, C. (1998). Motion-based Recognition of Pedestrians. In *Procs. IEEE Intl. Conf. on Pattern Recognition, June* (pp. 1325–1330).

H.Vold, Mains, M., & Blough, J. (1997). Theoretical Foundations for High Performance Order Tracking with the Vold-Kalmen Tracking Filter. *SAE Paper No. 972007*, (pp. 1083–1088).

Ikizle, N. & Forsyth, D. (2008). Searching for Complex Human Activities with No Visual Examples. *Int. J. Computer Vision, 80(3)*, (pp. 337–357).

Isard, M. & Blake, B. (1998). CONDENSATION - Conditional Density Propagation for Visual Tracking. *International Journal of Computer Vision*, 29(1), 5–28.

Ivanov, Y., Stauffer, C., Bobick, A., & Grimson, W. (1999). Video Surveillance of Interactions. *In 2nd International Workshop on Visual Surveillance, Fort Collins Colorado*, (pp. 82–89).

Iwasawa, S., Ebihara, K., Ohya, J., & Morishima, S. (1999). Real-TimeEstimation of Human Body Posture from Monocular Thermal Images. In *Conference on Computer Vision and Patern Recognition 73 (3)* (pp. 428–440).

Jafari, R., Li, W., Bajcsy, R., Glaser, S., & Sastry, S. (2007). Physical Activity Monitoring for Assisted Living At Home. *4th International Workshop on Wearable and Implantable Body Sensor Networks*, (pp. 213–219).

Jatoba, L., Grossmann, U., Kunze, C., Ottenbacher, J., & Stork, W. (2008). Context-aware Mobile Health Monitoring: Evaluation Of Different Pattern Recognition Methods for Classification of Physical Activity. In *30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 5250–5253).

Jhuang, H., Serre, T., L.Wolf, & Poggio, T. (2007). A Biologically Inspired System for Action Recognition. In *In Proceedings of the International Conference On Computer Vision (ICCV'07), Rio de Janeiro Brazil October* (pp. 1–8).

Ju, S., Black, M., & Yacoob, Y. (1996). Cardboard People: A Parameterized Model of Articulated Motion. In *2nd Int. Conf. on Automatic Face- and Gesture-Recognition. Killington Vermont* (pp. 38–44).

Kale, A., Rajagopalan, A., Cuntoor, N., & Krueger, V. (2002). Human Identification Using Gait. *Proc. Int. Conf. on Automatic Face and Gesture Recognition Washington DC USA.*, (pp. 137–142).

Kass, M., Witkin, A., & Snakes:, D. T. (1988). Active Contour Models. *International Journal of Computer Vision , Vol. 2 , No. 3*, (pp. 321–331).

Kehl, R., Bray, M., & L.VanGool (2005). Full Body Tracking from Multiple Views Using Stochastic Sampling. *Proc.IEEE Computer Vision and Pattern Recognition. Vol. 2*, (pp. 129–136).

Kim, K., Jung, K., Park, S., & Joon, H. (2002). Support Vector Machines for for Texture Classification. *IEEE Trans. on Pattern Analysis and Machine Intelligence 24(11)*, (pp. 1542–1550).

Krahnstover, N., Yeasin, M., & Sharma, R. (2001). Towards a Unified Framework for Tracking and Analysis of Human Motion. *IEEE Workshop on Detection and Recognition Events inVideo Vancouver Canada July*, (pp. 47–54).

Kruppa, H., M.Castrillán, & Schiele, B. (2003). Fast and Robust Face Finding Via Local Context. *VS-PETS Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking Surveillance Nice France*, (pp. 157–164).

Kusakunniran, W., Wu, Q., Zhang, J., & Li, H. (2010). Multi-view Gait Recognition Based on Motion Regression Using Multilayer Perceptron. *20th International Conference on Pattern Recognition (ICPR), August 2186-2189.*

Kushwaha, A., Sharma, C., Khare, M., Srivastava, R., & Khare, A. (2012). Automatic multiple human detection and tracking for visual surveillance system. In *Informatics, Electronics Vision (ICIEV), 2012 International Conference on* (pp. 326–331).

Lan, X. & Huttenlocher, D. P. (2005). Beyond Trees: Common-factor Models for 2D Human Pose Recovery. *In Proc. IEEE Conf. on ICCV*, (pp. 470–477).

Laptev, I., Caputo, B., Schüldt, C., & Lindeberg, T. (2007). Local Velocity-adapted Motion Events for Spatio-temporal Recognition. *Computer Vision and Image Understing (CVIU) 108 (3)*, (pp. 207–229).

Laptev, I. & Pérez, P. (2007). Retrieving Actions in Movies. In *in: Proceedings of the International Conference On Computer Vision (ICCV'07), Rio de Janeiro and Brazil October* (pp. 1–8).

Lawrence, L., Giles, & Recognition:, T. y. B. . F. (1997). A Convolutional Neural Network Approach. *IEEE Trans. Pattern Anal 8(1)*, (pp. 98–113).

Leigh, W., Purvis, R., & Ragusa, J. (2002). Forecasting the Nyse Composite Index with Technical Analysis, Pattern Recognizer Neural 36 Networks, and Genetic Algorithm: A Case Study in Romantic Decision Support. *Decis. Support Syst., 32*, (pp. 361–377).

Leo, M., D'Orazio, T., Gnoni, I., Spagnolo, P., & Distante, A. (2004). Complex Human Activity Recognition for Monitoring Wide Outdoor Environments. *Proc. Int. Conf. On Pattern Recognition, Cambridge UK. Vol 4.*, (pp. 913–916).

Leo, M., Spagnolo, P., Attolico, G., & Distante, A. (2003). Shape Based People Detection for Visual Surveillance Systems. In J. Kittler & M. Nixon (Eds.), *Audio- and Video-Based Biometric Person Authentication*, volume 2688 of *Lecture Notes in Computer Science* (pp. 285–293). Springer Berlin Heidelberg.

Lin, H., Kao, Y., Yang, F., & Wang, P. (2006). Content-based Image Retrieval Trained by Adaboost for Mobile Application. *IJPRAI, 20(4)*, (pp. 525–542).

Lin, J., Wu, Y., & Huang, T. S. (2000). Modeling Human Hard Constraint. *In Proc Workshop on Human Motion*, (pp. 121–126).

Lipton, A. J., Fujiyoshi, H., & Patil, R. S. (1998). Moving Target Classification and Tracking from Real-time Video. In *Proceedings of the 4th IEEE Workshop on Applications of Computer Vision (WACV'98)*, WACV '98 (pp. 129–136). Washington, DC, USA: IEEE Computer Society.

List, T., Bins, J., Fisher, R., & Tweed, D. (2005). A plug-and-play architecture for cognitive video stream analysis. In *Computer Architecture for Machine Perception, 2005. CAMP 2005. Proceedings. Seventh International Workshop on* (pp. 67–72).

Lou, J., Liu, Q., Tan, T., & Hu, W. (2002). Semantic Interpretation of Object Activities in a Surveillance System. In *Proc. International Conference on Pattern Recognition* (pp. 777–780).

Lu, C. & Ferrier, N. (2004). Repetitive Motion Analysis: Segmentation And Event Classification. *IEEE Trans. Pattern Analysis and Machine Intelligence 26(2).*, (pp. 258–263).

Luo, Y., Wu, T. W., & Hwang, J. N. (2003). Object-based Analysis and Interpretation of Human Motion in Sports Video Sequences by Dynamic Bayesian Networks. *Computer Vision and Image Understing 92.*, (pp. 196–216).

L.Wang, Ning, H., Tan, T., & Hu, W. (2003). Fusion of Static and Dynamic Body Biometrics for Gait Recognition. *International Conference on Computer Vision, Nice France. vol. 2*, (pp. 1449–1454).

Lyons, D. & Pelletier, D. (2000). A Line Scan Computer Vision Algorithm for Identifying Human Body Features. *in Lecture Notes in AI 1739 A. Braffart and et al. Editor. Springer-Verlag.*, (pp. 87–99).

Maimon, O. & Rokac, L. (2004). Ensemble of decision trees for mining manufacturing data sets. *In Machine Engineering 4 32-57*, (pp. 1–2).

Mangiameli, P., West, D., & Rampal, R. (2004). Model Selection for Medical Diagnosis Decision Support Systems. *Decis. Support Syst. 36*, (pp. 247–259).

Marr, D. & Nishihara, H. K. (1975). Spatial Disposition of Axes in a Generalized Cylinder Representation of Objects That Do Not Encompass the Viewer. *Massachusetts Institute of Technology and Artificia Intelligence Laboratory.*

Marr, D. & Nishihara, H. K. (1978). Representation and Recognition of the Spatial Organization of Three-dimensional Shapes. *Proc. Roy. Soc. London. Ser. B. , vol. 200 , no. 1140*, (pp. 269–294).

Martinez-Tomas, R., Rincón, M., Bachiller, M., & Mira, J. (2008). On the Correspondence Between Objects and Events for the Diagnosis of Situations in Visual Surveillance Tasks. *Pattern Recognition Letters.Elsevier 29(8)*, (pp. 1117–1135).

Matsuyama, Y. (2003). The alpha;-EM algorithm: surrogate likelihood maximization using alpha;-logarithmic information measures. *Information Theory, IEEE Transactions on*, 49(3), 692–706.

Matsuyama, Y. (2011). Hidden Markov model estimation based on alpha-EM algorithm: Discrete and continuous alpha-HMMs. In *Neural Networks (IJCNN), The 2011 International Joint Conference on* (pp. 808–816).

Maurer, U., Smailagic, A., Siewiorek, D. P., & Deisher, M. (2006). Activity Recognition and Monitoring Using Multiple Sensors on Different Body Positions. In *Proceedings of the International Workshop on Wearable and Implantable Body Sensor Networks* (pp. 113–116).

McInerney, T. & Terzopoulos, D. (1995). Topologically Adaptable Snakes. In *Computer Vision 1995. Proceedings Fifth International Conference.* (pp. 840–845).

Medioni, G., Cohen, I., Brémond, F., Hongeng, S., & Nevatia, R. (2001). Event Detection and Analysis from Video Streams. *in IEEE Transactions on Pattern Analysis and Machine Intelligence 23(8)*, (pp. 873–889).

Meeds, E., Ross, D., Zemel, R., & Roweis, S. (2008). Learning Stick-figure Models Using Nonparametric Bayesian Priors Over Trees. In *Proceedings of the EEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1–8).

Merkwirth, C., Mauser, H., Gasch, T. S., Roche, O., Stahl, M., & Lengauer, T. (2004). Ensemble Methods for Classification in Cheminformatics. *J Chem Inf Comput Sci 44(6)*, (pp. 1971–1978).

Minnen, D., Westeyn, T., Ashbrook, D., Presti, P., & Starner, T. (2007). Recognizing Soldier Activities in the Field. In *4th International Workshop on Wearable and Implantable Body Sensor Networks vol. 13 , Springer Berlin Heidelberg* (pp. 236–241).

Montemayor, A. P. J. & Sanchez, A. (2005). Recognizing Activities in the Field. In *The 32nd International Conference on Computer Graphics and Interactive Techniques Los Angeles, CA (USA), 31 July-4 August, 2005.*

Murphy, K. (2005). Hidden Markov Model (HMM) Toolbox for Matlab. http://www.cs.ubc.ca/ murphyk/Software/HMM/hmm.html.

Nair, V. & Clark, J. J. (2002). Automated Visual Surveillance Using Hidden Markov Models. In *International Conference on Vision Interface* (pp. 88–93).

Nakazawa, A., Kato, H., & Inokuchi, S. (1998). Human Tracking Using Dis- Luo and Distributed Video Systems. In *International Conference on Pattern Recognition , Volume 1 593.*

Nixon, M., Tan, T., & Chellappa, R. (2006). Human Identification Based on Gait. *International series on Biometrics, Springer*, (pp. 135–149).

Niyogi, S. A. & Adelson, E. H. (1994). Analyzing Gait with Spatiotemporal Surfaces. *Proc.IEEE Workshop on Nonrigid and Articulated Motion*, (pp. 64–69).

Nowozin, S., Bakır, G., & Tsuda, K. (2007). Discriminative Subsequence Mining for Action Classification. In *Proceedings of the International Conference On Computer Vision (ICCV'07), Rio de Janeiro Brazil October* (pp. 1–8).

Ogden, B. & Dautenhahn, K. (2001). Embedding Robotic Agents in the Social Environment. *Towards Intelligent Mobile Robots TIMR 01 , vol 3*, (pp. 397–428).

Oliver, N. M., Rosario, B., & Pentland, A. P. (2000). A Bayesian Computer Vision System for Modeling Human Interactions. *IEEE Trans. Pattern Analysis and Machine Intelligence 22(8)*, (pp. 831–843).

Ong, E. & Gong, S. (1999). A Dynamic Human Model Using Hybrid 2D-3D Representations in Hierarchical PCA Space. *Westfield College London*, (pp. 33–42).

Ozer, I. B. & Wolf, W. H. (2002). A Hierarchical Human Detection System in Compressed Domains. *IEEE Transactions on Multimedia 4(2)*, (pp. 283–300).

Pan, S. (2000). 3D Motion Estimation of Human by Genetic Algorithm. *15th International Conference on Pattern Recognition - Volume 1*, (pp. 11–59).

Parameswaran, V. & Chellappa, R. (2003). View Invariants for Human Action Recognition. *Computer Vision and Pattern Recognition Madison Wisconsin. Vol 2*, (pp. 613–619).

Pascual, J., Neucimar, J., & Barros, R. (2006). Tracking Soccer Players Aiming Their Kinematical Motion Analysis. *Computer Vision and Image Understing Vol. 101 , No. 2. February*, (pp. 122–135).

Pham, N. & Abdelzaher, T. (2008). Robust Dynamic Human Activity Recognition Based on Relative Energy Allocation. In *Distributed Computing in Sensor Systems , vol. 5067 of Lecture Notes in Computer Science Springer Berlin/ Heidelberg* (pp. 525–530).

Pomares, J., Torres, F., & Gil, P. (2002). 2-D Visual Servoing with Integration of Multiple Predictions of Movement Based on Kalman Filter. *IFAC 2002 15thWorld Congress, Barcelona*, (pp. 12–16).

Poppe, R. (2010). A Survey on Vision-based Human Action Recognition. *Image & Vision Computing 28(6)*, (pp. 976–990).

Quinlan, J. (1993). C4 5: programs for machine learning. *Morgan Kaufmann series in machine learning and Morgan Kaufmann Publishers*.

Rahimi, A., Dunagan, B., & Darrell, T. (2004). Tracking People with a Sparse Network of Bearing Sensors. *European Conference on Computer Vision (ECCV)*, (pp. 507–518).

Ramanan, D. & Forsyth, D. A. (2003). Finding and Tracking People from the Bottom Up. *CVPR (2)*, (pp. 467–474).

Ramírez, J., Yélamos, P., Górriz, J. M., & Segura, J. C. (2006). SVM-based Speech Endpoint Detection Using Contextual Speech Features. *IEE Electronic Letters ISSN:Vol. 42(7) 65-66*, (pp. 65–66).

Randell, C. & Muller, H. (2000). Context Awareness by Analysing Accelerometer Data. In *The Fourth International Symposium on Wearable Computers* (pp. 175–176).

Rao, C., Yilmaz, A., & Shah, M. (2002). View-Invariant Representation and Recognition of Actions. *En Journal of Computer Vision and 50*, (pp. 203–226).

Rincón, M., Carmona, E. J., Bachiller, M., & Folgado, E. (2007). Segmentation of Moving Objects with Information Feedback Between Description Levels. In *Proceedings of the 2nd international work-conference on Nature Inspired Problem-Solving Methods in Knowledge Engineering: Interplay Between Natural and Artificial Computation, Part II*, IWINAC '07 (pp. 171–181). Berlin, Heidelberg: Springer-Verlag.

Rincón, M., Folgado, E., Carmona, E., & Bachiller, M. (2006). Patente: Procedimiento para describir el comportamiento geométrico de humanos en una escena captada por un sistema de visión artificial, basado en un modelo de bloques y, en especial, orientado a la tarea de video-vigilancia. Request Number of Patent: Spain-P200603272, 27 Dic 2006. Entidad titular: UNED.

Rittscher, J., Blake, A., & Roberts, S. J. (2002). Towards the Automatic Analysis of Complex Human Body Motions. *Image and Vision Computing", 20 (12).*, (pp. 905–911).

Roberts, T. J., McKenna, S. J., & Ricketts, I. W. (2004). Human Pose Estimation Using Learnt Probabilistic Region Similarities and Partial Configurations. *Computer Vision - ECCV 2004 , 8th European Conference on Computer Vision, Prague Czech Republic*, (pp. 291–303).

Robertson, N. & Reid, I. (2005). Behaviour Understanding in Video: A Combined Method. *Proc. Int. Conf. on Computer Vision (ICCV05). Vol 1.*, (pp. 808–815).

Roesser, R. P. (1975). A Discrete State-space Model for Linear Image Processing. *IEEE Transactions on Automatic Control AC-20*, (pp. 1–10).

Sarkar, S., Phillips, P., Liu, Z., Vega, I., Grother, P., & Bowyer, K. (2005). The humanID gait challenge problem: data sets, performance, and analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(2), 162–177.

Sato, K. & Aggarwal, J. K. (2001). Tracking and Recognizing Two-person Interactions in Outdoor Image Sequences. *Workshop on Multi-Object Tracking, Vancouver Canada*, (pp. 87–94).

Schmidbauer, O. (1989). Robust Statistic Modelling of Systematic Variabilities in Continuous Speech Incorporating Acoustic-articulatory Relations. *In ICASSP-89*, (pp. 616–619).

Schüldt, C., Laptev, I., & Caputo, B. (2004). Recognizing Human Actions: A Local SVM Approach. In *Proceedings of the International Conference on Pattern Recognition (ICPR'04), 2004 , vol. 3 , Cambridge United Kingdom* (pp. 32–36).

Sidenbladh, H., Black, M., & Fleet, D. (2000). Stochastic Tracking of 3D Human Figures Using 2D Image Motion. In *Proc.ECCV* (pp. 702–718).

Siebel, N. & Maybank, S. (2004). The ADVISOR Visual Surveillance System. *Cognitive Systems Group, Institute of Computer Science Applied Mathematics School of Computer Science Information Systems Birkbeck College.*

Siebel, N., Maybank, S., Clabian, M., Smutny, V., G.Stanke, A.Shats, J.-d.-S., & Verschae, R. (2004). The ADVISOR Visual Surveillance System. In *Proceedings of the ECCV 2004 workshop "Applications of Computer Vision" (ACV'04), Prague Czech RepublicMay 2004 , ISBN 80-01-02977-8 , 103-111* (pp. 103–111).

Sminchisescu, C., A.Kanaujia, Z.Li, & Metaxas, D. (2004). *Learning to Reconstruct 3D Human Motion from Bayesian Mixture of Experts, a Probabilistic Discriminative Approach.* Technical report. CSRG-502 , University of Toronto.

Sminchisescu, C. & B.Triggs (2003). Kinematic Jump Processes for Monocular 3D Human Tracking. *IEEE International Conference on Computer Vision and Pattern Recognition CVPR (1)*, (pp. 69–76).

Smisek, J., Jancosek, M., & Pajdla, T. (2011). 3D with Kinect. *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference.*, (pp. 1154–1160).

Smith, P., Lobo, N. d. V., & Shah, M. (2005). Temporal Boost for Event Recognition. In *Proceedings of the International Conference On Computer Vision (ICCV'05), vol. 1 , Beijing China October* (pp. 733–740).

Sokolova, M. S.-C. J. C. J. & Fernández-Caballero, A. (2013). Fuzzy model for human fall detection in infrared video. In *Journal of Intelligent and Fuzzy Systems. Special Issue: Recent Advances in Intelligent and Fuzzy Systems* (pp. 1064–1246).

Soltani, F., Eskandari, F., & S.Golestan (2012). Developing a Gesture-Based Game for Deaf/Mute People Using Microsoft Kinect. *Complex Intelligent and Software Intensive Systems (CISIS) Sixth International Conference*, (pp. 491–495).

Spagnolo, P., Leo, M., Attolico, G., & Distante, A. (2003). Posture recognition in visual surveillance of archeological sites. In *Intelligent Robots and Systems, 2003. (IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on*, volume 2 (pp. 1542–1547 vol.2).

Starner, T. & Pentland, A. (1995). Real-time American Sign Language recognition from video using hidden Markov models. In *Computer Vision, 1995. Proceedings., International Symposium on* (pp. 265–270).

Stauffer, C. & Grimson, W. (2000). Learning Patterns of Activity Using Real-time Tracking. *IEEE Trans. Pattern Analysis and Machine Intelligence 22(8)*, (pp. 747–757).

Sung, U. & Cohen, I. (2004). 3D Hand Reconstruction from a Monocular View. *International Conference on Pattern Recognition. Cambridge United Kingdom vol. 3*, (pp. 310–313).

Tany, H. (2004). Surveillance of Pedestrians on Crosswalk by Camera. In *Transactions of the Institute of Electrical Engineers of Japan and Part-C.* (pp. 798–804).

Thalmann, M. N. & Seo, H. (2004). Data-driven Approaches to Digital Human Modeling. *Proc. 2nd International Symposium on 3D Data Processing and Visualization Transmission Thessalonica Greece IEEE Computer Society Press. 380-387.*

Thonnat, M. & Rota, N. (1999). Image Understanding for Visual Surveillance Applications. In *Third International Workshop on Cooperative Distributed Vision, Kyoto Japan November* (pp. 51–82).

Thonnat, M. & Rota, N. (2000). Video Sequence Interpretation for Visual Surveillance. In *INRIA Sophia Antipolis, France. In Visual Surveillance 2000. Proceedings. Third IEEE International Workshop on* (pp. 59–68).

Trinh, H., Fan, Q., Pan, J., Gabbur, P., Miyazawa, S., & Pankanti, S. (2011). Detecting human activities in retail surveillance using hierarchical finite state machine. In *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on* (pp. 1337–1340).

Vecchio, D., Murray, R. M., & Perona, P. (2003). Decomposition of Human Motiom Into Dynamics-based Primitives with Application to Drawing Tasks. *Automatica. vol. 39*, (pp. 2085–2098).

Vinh, L., Lee, S., Le, H., Ngo, H., Kim, H., Han, M., & Lee, Y. K. (2011). Semi-markov Conditional Random Fields for Accelerometer-based Activity Recognition. *Applied Intelligence , vol 35*, (pp. 226–241).

Wang, L., Tan, T., Ning, H., & Hu, W. (2003). Silhouette Analysis-Based Gait Recognition for Human Identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence 25(12)*, (pp. 1501–1518).

Welch, L. R. (2003). Hidden Markov Models and the Baum-Welch Algorithm. *IEEE Information Theory Society Newsletter*, 53(4).

Wilson, A. D. & Bobick, A. F. (1999). Parametric Hidden Markov Models for Gesture Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence 21(9)*, (pp. 884–900).

Wren, C. R., Azarbayejani, A., Darrell, T., & Pentland, A. (1997a). Pfinder: Real-time Tracking of the Human Body. *IEEE Trans.on Pattern Analysis and Machine Intelligence. 19(7)*, (pp. 780–785).

Wren, C. R., Sparacino, F., Azarbayejani, A., Darrell, T., Starner, T., Chao, K. A., Hlavac, M., Russell, K., & Pentland, A. (1997b). PerceptiveSpaces for Peformance and Entertainment: Untethered InteractionUsing Computer Vision and Audition. *In Applied Artificial Intelligence 11(4)*, (pp. 267–284).

Wu, Y., Lin, J., & Huang, T. S. (2001). Capturing Natural Hand Articulation. In *Proc.IEEE Int. Conf. on Computer Vision (ICCV'01), Vancouver Canada vol. 2* (pp. 426–432).

Yam, C., Nixon, M., & Carter, J. (2002). On the Relationship of Human Walking and Running: Automatic Person Identification by Gait. *International Conference on Pattern Recognition, Quebec Canada. vol. 1.*, (pp. 287–290).

Yi, H., Rajan, D., & Chia, L. T. (2004). *A New Motion Histogram to Index Motion Content in Video Segments*. *Pattern Recognition Letters , vol. 26*, (pp. 1221–1231).

Yu, X. & Yang, S. X. (2005). A Study of Motion Recognition from Video Sequences. *Computing and Visualization in Science vol.8*, (pp. 19–25).

Zhang, F., Wang, Y., & Zhang, Z. (2011). View-invariant action recognition in surveillance videos. In *Pattern Recognition (ACPR), 2011 First Asian Conference on* (pp. 580–583).

Zhang, J., Collins, R., & Liu, Y. (2004). Representation and Matching of Articulated Shapes. *In Proceedings of the 2004 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'04), Vol. 2*, (pp. 342–349).

Zhang, T. & Freedman, D. (2005). Improving Performance of Distribution Tracking Through Background Mis- Match. *Pattern Analysis and Machine Intelligence IEEE Transactions on 27 (2)*, (pp. 282–287).

Zhao, L. & Thorpe, C. (2000). Stereo and Neural Network-based Pedestrian Detection. *IEEE Transactions on Intelligent TransportationSystems 1(3)*, (pp. 148–154).

Zhu, C. & Sheng, W. (2009). Human Daily Activity Recognition in Robot Assisted Living Using Multi-sensor Fusion. *In IEEE International Conference on Robotics and Automation*, (pp. 2154–2159).

Capítulo 7

Anexo I

7.1. Herramienta WEKA

La herramienta Weka, Group (2012), ofrece 4 modos de funcionamiento (Applications):

- Explorer. Interfaz básico para usar el conjunto de algoritmos que ofrece WEKA (clasificación, clustering, reglas de asociación, selección de atributos y visualización)
- Experimenter. Interfaz gráfico para automatizar baterías de experimentos
- Knowledge Flow. Interfaz gráfico para diseñar flujos y procesos donde se combinen varios componentes para conformar aplicaciones complejas.
- Simple CLI. Acceso los algoritmos de WEKA desde un interfaz de línea de comandos.

El modo interfaz más común en el que se utiliza WEKA es en el de explorer. Por ello, a continuación se explica brevemente el mismo. Las funcionalidades del interfaz Explorer se organizan en 6 pestañas:

- Preprocess.
  Carga de los datasets a emplear y procesamiento previo a la aplicación de los algoritmos de aprendizaje.
  - Permite cargar datos desde ficheros ARFF, fichero CSV y bases de datos (mediante JDBC).
  - Permite aplicar distintos filtros sobre los datos cargados: selección de atributos, selección de instancias, modificación/transformación de atributos (conversión numérico-nominal, conversión texto-vector_numérico, etc).

- Permite seleccionar manualmente los atributos a emplear y analizar su distribución.

- Classify
  Interfaz de experimentación con algoritmos de clasificación.

  - Permite seleccionar un clasificador (botón [Choose]) y configurar sus parámetros (pulsando sobre el nombre del algoritmo).
  - Permite especificar el método de evaluación (Test options).
    - Training set. Usa los mismos datos para el entrenamiento y para la evaluación.
    - Supplied test set. Usa un nuevo fichero ARFF para realizar la evaluación.
    - Cross validation. Divide los datos disponibles en k grupos y realiza k tandas de entrenamiento-evaluación diferentes, usando $k-1$ grupos para entrenar y el resto para la validación.
    - Percentage split. Divide los datos disponibles en un grupo de entrenamiento (basándose en el porcentaje indicado) y en otro grupo de evaluación.
  - Permite establecer el atributo clase sobre el que realizar el aprendizaje.
  - Muestra los resultados del proceso entrenamiento-evaluación.

- Cluster.
  Interfaz de experimentación con algoritmos de clustering.

  - Associate. Interfaz de experimentación con algoritmos para aprendizaje de reglas de asociación.
  - Select attributes. Interfaz de experimentación con algoritmos de selección de atributos.
  - Visualize. Interfaz para la visualización de datasets, relaciones entre atributos, etc.

7.1.1. Algoritmos utilizados

A continuación, se presentan algunos de los algoritmos implementados por WEKA y que se han utilizado en esta tesis. Después, se explicarán las opciones más relevantes de los algoritmos utilizados.

7.1.1.1. Algoritmo J48

El algoritmo J48 de WEKA, es una implementación del algoritmo C4.5 Quinlan (1993),es uno de los algoritmos más utilizados en el ámbito de los árboles de clasificación. J48 es la implementación de una versión mejorada del mismo llamada C4.5 revisión 8. Estos algoritmos permiten formar árboles de decisión para clasificar instancias.

Dentro de las opciones que J4.8 soporta están:

- La poda de árboles.
- La especificación de factores de confianza para la poda.
- La especificación de un mínimo de instancias en las hojas.
- La poda de árboles con error reducido.
- La especificación del número de datos en podas con error reducido.
- El uso de particiones binarias en atributos nominales.

El parámetro más importante que se debe tener en cuenta es el factor de confianza FC para la poda "confidence factor", que influye en el tamaño y capacidad de predicción del árbol construido, Cuanto menor es el factor menor cantidad de poda. El valor por defecto es del 25 %.

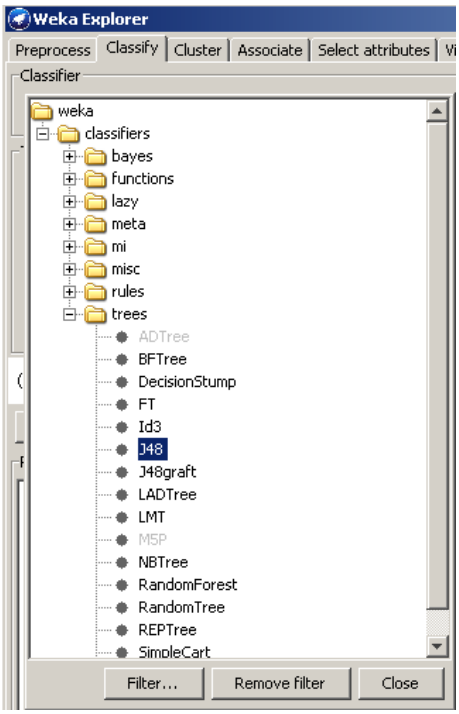


Figura 7.1: Selección del algoritmo de J48 en Weka.

La implementación de J48 en WEKA se encuentra en el apartado "TREES" (fig. 7.1). Las opciones más relevantes para esta tesis son las siguientes:

- MinNumObj (2). Número mínimo de instancias por hoja.
- SaveInstanceData (False). Una vez finalizada la creación del árbol de decisión se eliminan todas las instancias que se clasifican en cada nodo, que hasta el momento se mantenían almacenadas.
- BinarySplits (False). Con los atributos nominales también no se divide (por defecto) cada nodo en dos ramas.
- Unpruned (False). En caso de no activar la opción se realiza la poda del árbol.
- SubtreeRaising (True). Permite realizar la poda con el proceso subtree raising (se aplica según se va construyendo el árbol).
- ConfidenceFactor (0.25). Factor de confianza para la poda del árbol.
- ReducedErrorPruning (False). Si se activa esta opción, el proceso de poda no es el propio de C4.5, sino que el conjunto de ejemplos se divide en un subconjunto de entrenamiento y otro de test, de los cuales el último servirá para estimar el error para la poda.
- NumFolds (3). Define el número de subconjuntos en que hay que dividir el conjunto de ejemplos para, el último de ellos, emplearlo como conjunto de test si se activa la opción reducedErrorPruning.
- UseLaplace (False). Si se activa esta opción, cuando se intenta predecir la probabilidad de que una instancia pertenezca a una clase, se emplea el suavizado de Laplace.

7.1.1.2. Algoritmo de Bagging

La implementación de Bagging en WEKA se encuentra en el apartado META (fig. 7.2). Las opciones más relevantes para esta tesis son las siguientes:

- BagSizePercent. Indica el porcentaje de casos seleccionados para generar las muestras boostrap.
- CalcOutOfBag. Indica si fuera de la bolsa el error se calcula.
- Classifier. Es el clasificador base que es utilizado.
- NumIterations. El número de iteraciones a realizar.
- Seed. Es el número aleatorio de semillas que son utilizados.

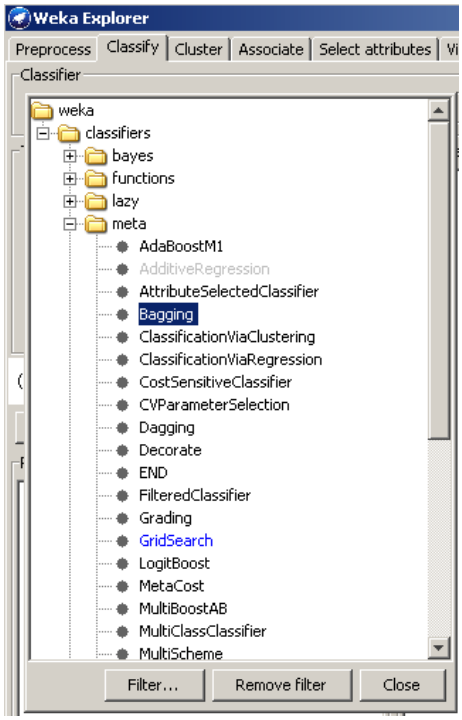


Figura 7.2: Selección de los Metabuscadores.

7.1.1.3. Algoritmo de boosting

La implementación de AdaBOOST en WEKA se encuentra en el apartado META.

Opciones relevantes para esta tesis:

- Classifier. Es el clasificador base que es utilizado.
- UseResampling. Indica si el boosting utiliza un peso de las instancias y una posterior actualización de los mismos en cada iteración, o simplemente se usa un remuestreo de las instancias.
- WeightThreshold. Indica el porcentaje de casos seleccionados para generar las muestras boostrap.
- NumIterations. Indica el número de iteraciones del boosting.

7.1.1.4. Algoritmo de Stacking

La implementación de Stacking en WEKA se encuentra en el apartado META.
Opciones más relevantes para esta tesis:

- Classifier Es el clasificador base que es utilizado.

- NumFolds Define el número de subconjuntos para el cross-validation (dado un número n se divide los datos en n partes y, por cada parte, se construye el clasificador con las $n-1$ partes restantes y se prueba con esa. Así por cada una de las n particiones).

7.1.1.5. Algoritmo MLP

La implementación del Perceptrón Multicapa (MLP) en WEKA se encuentra en el apartado FUNCTIONS, (fig. 7.3)

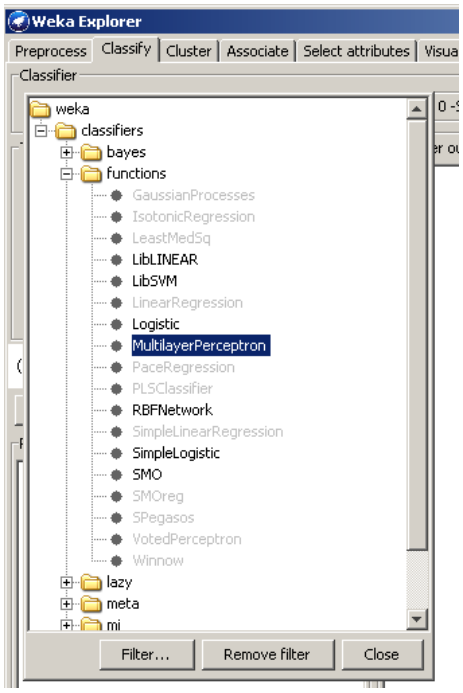


Figura 7.3: Selección del perceptrón multicapa (MLP) en Weka.

Opciones relevantes para esta tesis:

- HiddenLayers. Número de neuronas en la capa oculta.

- LearningRate. Factor de actualización de pesos cuando han de ser modificados, un valor mayor converge más rápido pero puede provocar oscilaciones.

- ValidationThreshold. Determina cuando dar por acabado el test de validación, cuántas veces en una fila el error puede empeorar antes de terminar el entrenamiento.

7.1.1.6. Algoritmo SVM

La implementación del SVM en WEKA se encuentra en el apartado denominado, FUNCTIONS, (fig. 7.4).

Opciones relevantes para esta tesis:

- Epsilon. Valor de épsilon.
- C. Complejidad
- ToleranceParameter. Tolerancia
- FilterType. Normalización o no.
- NumFolds. Número de subconjuntos para el cross-validation
- Kernel. Tipo de Kernel.

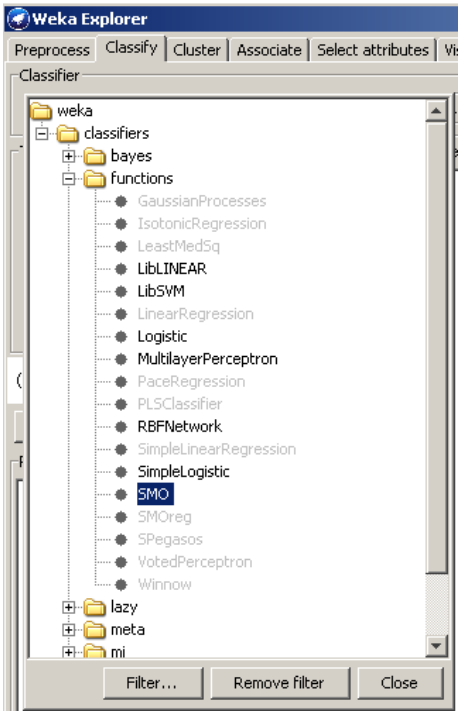


Figura 7.4: Selección de la máquina de soporte vectorial (SVM) en Weka.

Cuando se realiza la evaluación de los datos según el método de evaluación escogido se presentan los siguientes datos estadísticos: instancias bien clasificadas (Bien %), instancias mal clasificadas (Mal %), estadísticas Kappa (Ks), error medio absoluto (Mae), raíz del error cuadrático medio (Rmse), error relativo (Rae %), raíz del error cuadrático relativo (Rrse %).

Capítulo 8

Anexo II

8.1. Paquete Murphy Toolbox (HMM)

Este paquete se utiliza para proporcionar la funcionalidad básica para realizar clasificaciones utilizaando HMMs tanto en tiempos discretos como continuo, así como otras funcionalidades estadísticas necesarias para trabajar con los HMMs. Se presenta una breve descripción de las características de la herramienta y la funcionalidad ofrecida. Para consulta bibliográfica sobre HHM ver sección 2.4.2.

8.1.1. Funciones de entrenamiento en tiempo discreto

La función dhmm_em se utiliza para entrenar una modelo discreto con el algoritmo de EM (realmente utiliza el algoritmo Baum-Welch, que es una simplificación del EM). Esta función recibe como parámetros un modelo oculto de Markov discreto, formado por un vector con las probabilidades iniciales del modelo, una matriz de probabilidades de transición, una matriz de probabilidades de observación y un conjunto de datos de entrenamiento e intenta entrenar el modelo que recibe como parámetro, para mejorar la capacidad del mismo para reconocer dichas secuencias.

Como resultado, produce un modelo mejorado y una matriz logarítmica de probabilidades, que se puede utilizar para comprobar si entre sucesivas iteraciones del algoritmo se mejora la probabilidad de observar una determinada cadena de entrenamiento. Se trata de un algoritmo de búsqueda local, con cierta tendencia a devolver máximos locales como resultado, con lo cual debe tenerse en cuenta que los puntos desde los que se parte el entrenamiento son de gran importancia.

La llamada a la función de entrenamiento es:

dhmm_em(data, prior, transmat, obsmat, varargin)

Siendo los parámetros especificados:

- data: matriz formada por dos vectores con los datos observados y clasificados (caso de haber un único vector). De haber varios, se presentan tantas cell, cada una con los vectores con los datos observados.
- prior: π ; probabilidades iniciales.
- transmat: A; matriz de probabilidades de transición.
- obsmat: B; matriz de probabilidades de observación.
- varargin: argumentos opcionales:
  - 'maxIter': número máximo de iteraciones.

La salida de la llamada a la función:

[LL, prior, transmat, obsmat, nrIterations] = dhmm_em(data, prior, transmat, obsmat, varargin)

Devuelve como resultado los parámetros del HMM entrenado en base a dichas observaciones:

- prior : π; probabilidades iniciales.
- transmat:A; matriz de probabilidades de transición.
- obsmat: B; matriz de probabilidades de observación.
- nrIterations: numero de iteraciones que se han realizado para alcanzar el máximo.
- LL: probabilidad de reconocer la o las secuencias de entrenamiento.

8.1.2. Funciones de entrenamiento en tiempo continuo

El entrenamiento continuo es bastante similar al discreto, salvo que al tratarse de HMM continuos, en lugar de utilizarse una matriz de probabilidades de observación, se utilizan mezclas funciones gaussianas de densidad de probabilidad para representar las probabilidades de las variables ocultas.

Estas gaussianas vienen dadas por dos parámetros, la media y la varianza, de forma que la llamada a la función de entrenamiento continuo, mhmm_em, se hace tomando como parámetros los datos de entrenamiento, las matrices de probabilidad inicial y de transición y los valores con la media y varianza de la gaussiana.

Como resultado devuelve un modelo mejorado y una matriz logarítmica de probabilidades, que se puede utilizar para comprobar si entre sucesivas iteraciones del

algoritmo, mejora la probabilidad de observar una determinada cadena de entrenamiento.

Al igual que con su variante discreta, esta función está basada en el algoritmo EM simplificado (Baum-Welch), de forma que se trata de una función de búsqueda local, por lo que en caso de encontrar máximos locales al entrenar un HMM los devolverá como resultado válido. Es por tanto muy dependiente de los valores iniciales que reciba como parámetro.

Llamada a la función:

mhmm_em(data, prior, transmat, mu, sigma, varargin)

Siendo los parámetros especificados:

- data: matriz formada por dos vectores con los datos observados y clasificados (caso de haber un único vector). De haber varios, se presentan tantas cell, cada una con los vectores con los datos observados.
- prior : π, probabilidades iniciales.
- transmat: A, matriz de probabilidades de transición.
- mu; μ_{jm}, medias de las gaussianas.
- sigma:$\tilde{O}_{jm}$, covarianzas de las gaussianas.
- varargin: argumentos opcionales:
  - 'maxIter': número máximo de iteraciones

Salida:

[LL, prior, transmat, obsmat, nrIterations] = mhmm_em(data, prior, transmat, obsmat, varargin)

Devuelve como resultado los parámetros del HMM entrenado en base a dichas observaciones:

- prior : π, probabilidades iniciales.
- transmat: A, matriz de probabilidades de transición.
- mu: μ_{jm}, medias de las gaussianas.
- sigma:$\tilde{O}_{jm}$, covarianzas de las gaussianas.
- nrIterations: numero de iteraciones que se han realizado para alcanzar el máximo.
- LL: probabilidad de reconocer la o las secuencias de entrenamiento.

8.1.3. Funciones de clasificación en tiempo discreto

La función dhmm_logprob se puede utilizar para clasificar una secuencia dado un modelo, devolviendo como resultado la probabilidad logarítmica de que en un HMM se de un determinado conjunto de observaciones.

Se utiliza cuando se tienen varios HMM candidatos a generar una determinada secuencia y se quiere saber cuál de todos ellos es el que tiene más probabilidades de reconocerla.

Recibe como parámetros una secuencia de observaciones y las matrices de probabilidades, inicial, de transición y de observación, devolviendo la probabilidad de que el modelo dado haya generado dicha secuencia de observaciones.

Los datos se asume que pertenecen a un conjunto {1,2, . . . ,S} siendo S el tamaño de la matriz de transición recibido. No pueden contener '0'.

La llamada a la función tiene las siguientes entradas:

[loglik, errors] = dhmm_logprob(data, prior, transmat, obsmat)

- data: O, observaciones.
- prior: π, probabilidades iniciales.
- transmat: A, matriz de probabilidades de transición.
- obsmat: B, matriz de probabilidades de observación.

La llamada a la función devuelve :

[loglik, errors] = dhmm_logprob(data, prior, transmat, obsmat)

devuelve los siguientes datos:

- loglik: log probabilidad de observar la o las secuencias de datos.
- errors: errores en las secuencias

8.1.4. Funciones de clasificación en tiempo continuo

La función mhmm_logprob es muy similar a dhmm_logprob: se puede utilizar para clasificar una secuencia dado un modelo, devolviendo como resultado la probabilidad logarítmica de que en un HMM se dé un determinado conjunto de observaciones.

Al igual que su variante discreta, se puede utilizar cuando se tienen varios HMM candidatos a generar una determinada secuencia y se quiere saber cuál de todos ellos es el que tiene más probabilidades de reconocerla, o para entrenamientos y evaluación de modelos ocultos de Markov, al generar un coeficiente dados un modelo

y un conjunto de observaciones, permitiendo dibujar una curva de aprendizaje para una secuencia dada y sucesivos entrenamientos de un modelo.

Recibe como parámetros una secuencia de observaciones y las matrices de probabilidades, inicial y de transición, así como los valores que definen las gaussianas de emisión de probabilidades ocultas, las medias y varianzas.

Devuelve como resultado la probabilidad de que el modelo dado haya generado dicha secuencia de observaciones. Los datos se asume que pertenecen a un conjunto {1,2,...,S} , siendo S el tamaño de la matriz de transición recibido. No pueden contener '0'.

La llamada a la función tiene las siguientes entradas:
[loglik, errors] = mhmm_logprob(data, prior, transmat, mu, sigma)

- data: O, observaciones.
- prior : π, probabilidades iniciales.
- transmat: A, matriz de probabilidades de transición.
- mu: μ_{jm}, medias de las gaussianas.
- sigma:$\tilde{O}_{jm}$, covarianzas de las gaussianas.

La llamada a la función devuelve :
[loglik, errors] = mhmm_logprob(data, prior, transmat, mu, sigma)
devuelve los siguientes datos:

- loglik: log probabilidad de observar la o las secuencias de datos.
- errors: errores en las secuencias.

8.1.5. Funciones de generación de observaciones a tiempo discreto

Se puede obtener una secuencia de observaciones que incluya datos observados y las transiciones ocultas mediante la función dhmm_sample, que recibe como parámetros las matrices que definen un HMM discreto así como las longitudes de la secuencia y el número de secuencias a generar basados en estos parámetros.

Se puede utilizar para generar secuencias artificiales de pruebas que incluyen además de la observación las transiciones ocultas, de forma que facilite los entrenamientos o las evaluaciones de los HMM.

Recibe como parámetros las longitudes de secuencia y número de secuencias a generar, cuando haga falta generar múltiples secuencias de observación, así como las

matrices que definen el modelo de Markov: probabilidad inicial, de transición y de observación.

Como resultado, genera una secuencia de valores observados y una secuencia de los valores ocultos que han dado lugar a dichas observaciones.

La llamada a la función tiene las siguientes entradas:

[obs, hidden] = dhmm_sample(initial_prob, transmat, obsmat, numex, len)

- initial_prob: π, probabilidades iniciales.
- transmat: A, matriz de probabilidades de transición.
- obsmat:B, matriz de probabilidades de observación.
- numex : número de secuencias a generar
- len: longitud de las secuencias a generar

La llamada a la función devuelve:

[obs, hidden] = dhmm_sample(initial_prob, transmat, obsmat, numex, len)

devuelve los siguientes datos:

- obs: datos observador.
- hidden: datos reales.

Los resultados tendrán **numex** filas y **len** columnas.

8.1.6. Funciones de generación de observaciones a tiempo continuo

Al igual que con el modelo discreto, se puede obtener una secuencia de observaciones que incluya datos observados y las transiciones ocultas mediante la función mhmm_sample.

Se puede utilizar para generar secuencias artificiales de pruebas que incluyen además de la observación las transiciones ocultas, de forma que facilite los entrenamientos o las evaluaciones de los HMM.

Recibe como parámetros las longitudes de secuencia y número de secuencias a generar (cuando haga falta generar múltiples secuencias de observación), así como las matrices de probabilidad, inicial y de transición y los valores que definen las gaussianas de emisión de probabilidades ocultas, las medias y varianzas.

Como resultado, genera una secuencia de valores observados y una secuencia de los valores ocultos que han dado lugar a dichas observaciones.

La llamada a la función tiene las siguientes entradas:
[obs, hidden] = mhmm_sample(initial_prob, transmat, mu, sigma, numex,len)

- initial_prob: π, probabilidades iniciales.
- transmat: A, matriz de probabilidades de transición.
- mu:μ_{jm}, medias de las gaussianas.
- sigma: $\tilde{O}_{jm}$, covarianzas de las gaussianas.
- numex : número de secuencias a generar.
- len: longitud de las secuencias a generar.

La llamada a la función devuelve:
[obs, hidden] = mhmm_sample(initial_prob, transmat, obsmat, numex, len)
devuelve los siguientes datos:

- obs: datos observador.
- hidden: datos reales.

Los resultados tendrán numex filas y len columnas.

8.1.7. Funciones de generación de la secuencia más probable

La función viterbi_path devuelve el camino más probable generado, utilizando la función fwdbac para calcular las distintas probabilidades. La función viterbi_path, que dadas las matrices de probabilidad inicial y de transición así como una matriz de observaciones condicionadas que viene dada por las funciones, multinomial_prob en la variante discreta y mixgauss_prob en la continua. Multinomial_prob calcula las matrices de observación con una función de cálculo multinomial discreto y mixgauss_prob las calcula con una mezcla de gaussianas continua.

La llamada a la función viterbi toma valores distintos en base a si el modelo es continuo o discreto.

Modelo discreto

Calcula la secuencia de estados más probable que puede haber generado una observación,

path=viterbi(prior, transmat, obsmat, data)

- data: observaciones, O
- prior: π, probabilidades iniciales.

- transmat: A, matriz de probabilidades de transición.
- obsmat: B, matriz de probabilidades de observación.

Como resultado, se obtiene:

- path: secuencia de observaciones más probable.

Modelo continuo

path=viterbi(prior, transmat, mu, sigma, data)

- data: observaciones, O
- prior : π, probabilidades iniciales.
- transmat: A, matriz de probabilidades de transición.
- mu:μ_{jm}, medias de las gaussianas.
- sigma:$\tilde{O}_{jm}$, covarianzas de las gaussianas.
- varargin: vector opcional con los pesos de las gaussianas. De no introducirse, se toman los mismos pesos para todas las gaussianas.

Como resultado, se obtiene:

path: secuencia de observaciones más probable.