

**UNIVERSIDAD AUTONOMA DE MADRID**

**ESCUELA POLITECNICA SUPERIOR**



**PROYECTO FIN DE CARRERA**

**SEGUIMIENTO DE PERSONAS EN VÍDEO BASADO EN  
DETECCIÓN**

**Raúl Porras Martín**

**Junio 2014**



# **SEGUIMIENTO DE PERSONAS EN VÍDEO BASADO EN DETECCIÓN**

**AUTOR: Raúl Porras Martín**  
**TUTOR: José María Martínez Sánchez**

**Grupo VPULab**  
**Dpto. Tecnología Electrónica y de las Comunicaciones**  
**Escuela Politécnica Superior**  
**Universidad Autónoma de Madrid**  
**Junio 2014**

**“Trabajo parcialmente financiado por el gobierno español bajo el proyecto  
TEC2011-25995 (EventVideo)”**





## Palabras clave

---

*Frame*, imagen, detección de objetos, seguimiento de objetos, asociación, oclusión, Latent SVM, Filtro de Kalman, Filtro de Partículas.

## Resumen

---

El objetivo principal de este proyecto es crear un sistema de seguimiento de múltiples personas basado en la combinación de la información proporcionada por un detector de personas y un *tracker*. Se trata, en primer lugar, de tener un sistema base sobre el que aplicar mejoras en su módulo de asociación de identidades.

Después de analizar en detalle el estado del arte, se desarrolla un protocolo de evaluación para determinar el rendimiento del algoritmo a implementar. Para ello, se propone un *dataset* con diferentes niveles de complejidad y se seleccionan algunas de las métricas más utilizadas por la comunidad investigadora para medir el rendimiento del sistema de seguimiento.

Seguidamente, se seleccionan algoritmos de detección de personas y de seguimiento de objetos, se genera un sistema base de seguimiento de múltiples personas que utiliza un proceso de asociación de detecciones entre *frames* consecutivos y se evalúa el algoritmo implementado utilizando las métricas seleccionadas.

Una vez implementado el sistema básico, se añaden mejoras al sistema base en su módulo de asociación de identidades, se evalúa el nuevo algoritmo y se comparan los nuevos resultados con los resultados obtenidos para el algoritmo anterior para cada una de las combinaciones de detectores de personas y algoritmos de seguimiento seleccionados.



## Abstract

---

Main objective of this project is to create a multiple-person tracking system based on a combination of information provided by a person detector and a tracker. The first goal is to develop a system which can be used as a base where you can implement improvements in its identities association module.

After analysing in detail the state of the art, an assessment protocol is developed to evaluate the implemented algorithm's performance. In this way, a dataset with different levels of complexity is proposed, and the most frequent state of the art metrics are chosen to measure the performance of the tracking system.

Then, people detection and object tracking algorithms are chosen. A base system of multiple-person tracking is generated. It uses a matching process to make the association of detections between consecutive frames. The implemented algorithm is assessed using previously selected metrics.

Identities association module is improved after the basic system implementation has been finished. The new algorithm is assessed and the new results are compared with the results obtained for the previous algorithm. It is done for every combination of people detectors and tracking algorithms chosen.



## ***Agradecimientos***

*En primer lugar, quiero dar mi más sincero agradecimiento a mi tutor Chema por haberme dado la oportunidad de realizar este proyecto. Por su atención y constante apoyo, siempre dispuesto a atenderme con buen humor para resolver dudas y comentar soluciones que hiciesen el trabajo más gratificante.*

*También, a cada uno de los componentes del VPULab que de una forma u otra me han ayudado a realizar este trabajo. La verdad que en este laboratorio se respira muy buen ambiente y todo el mundo se encuentra abierto a ayudar en lo que se necesite.*

*Me gustaría dedicar unas líneas de estos agradecimientos a mi compañero Ángel, que desde la etapa de bachillerato me ha acompañado. Nombrados como la pareja de prácticas mítica Porrás-Mora, hemos compartido muchas memorias, prácticas y tardes de sufrimiento.*

*A mis amigos de siempre, Adri, Carlos, Dani, Sampa y Unai por cada uno de los momentos que hemos pasado juntos y los buenos ratos vividos.*

*A cada uno de mis compañeros de clase y amigos, David, Guille, Jorge, Juanma, Mario y Pencho por cada uno de los momentos que hemos compartido en estos años, por esos viajes míticos a Lisboa y Cancún, por esos Achtung con cien manos en un ordenador que nos permitían descansar un rato en aquellas tardes de intensivo de prácticas.*

*Pero sin lugar a dudas a quien más tengo que agradecer su apoyo constante, su confianza en mí en todo momento y su ayuda incondicional, es a mis padres y a mi hermana Marta que han sabido guiarme en todo momento por el buen camino.*

*A todos muchas gracias.*

*Raúl Porrás Martín  
Junio de 2014*



# INDICE DE CONTENIDOS

1	Introducción.....	1
1.1	Motivación.....	1
1.2	Objetivos.....	1
1.3	Organización de la memoria.....	2
2	Estado del arte .....	3
2.1	Introducción.....	3
2.2	Detección de personas .....	3
2.2.1	Arquitectura de los sistemas de detección de personas .....	3
2.2.2	Clasificación de los algoritmos de detección de personas.....	5
2.2.2.1	Métodos de detección de objetos .....	5
2.2.2.1	Modelo de persona.....	7
2.3	Seguimiento de objetos en vídeo .....	9
2.3.1	Características.....	9
2.3.1.1	Características de bajo nivel .....	10
2.3.1.2	Características de nivel medio .....	10
2.3.1.3	Características de alto nivel .....	11
2.3.2	Representación.....	11
2.3.3	Predicción de movimiento .....	12
2.3.3.1	Búsqueda exhaustiva .....	12
2.3.3.1	Búsqueda selectiva .....	13
2.3.4	Clasificación de los algoritmos de seguimiento .....	13
2.3.4.1	Trackers basados en características discretas .....	14
2.3.4.2	Trackers basados en contornos .....	14
2.3.4.3	Trackers híbridos .....	15
2.3.4.4	Trackers basados en regiones .....	15
2.4	Seguimiento de múltiples personas basado en detección .....	16
2.5	Evaluación .....	17
2.5.1	Datasets.....	17
2.5.1.1	PETS.....	17
2.5.1.2	CAVIAR.....	18
2.5.1.3	VISOR .....	20
2.5.1.4	iLIDS .....	21
2.5.2	Métricas .....	21
2.5.2.1	Notación .....	22
2.5.2.2	Evaluación PAP.....	23
2.5.2.3	Evaluación RAP .....	24
2.5.2.4	Evaluación RAS .....	25
2.6	Conclusiones.....	26
3	Protocolo de evaluación propuesto.....	29
3.1	Introducción.....	29
3.2	Metodología de evaluación objetiva.....	29
3.2.1	Algoritmo de seguimiento .....	29
3.2.2	Evaluación del rendimiento .....	30
3.3	Metodología de evaluación subjetiva .....	31
3.4	Dataset .....	32

3.5 Métrica.....	36
3.6 Conclusiones.....	38
4 Algoritmos seleccionados.....	39
4.1 Introducción.....	39
4.2 Detector Latent SVM.....	39
4.3 Filtro de Kalman.....	44
4.3.1 Algoritmo del Filtro de Kalman .....	44
4.3.2 Filtro de Kalman seleccionado .....	47
4.4 Filtro de Partículas.....	48
4.4.1 Algoritmo del Filtro de Partículas .....	48
4.4.2 Filtro de Partículas seleccionado .....	49
4.5 Conclusiones.....	53
5 Sistema base y módulos comunes .....	55
5.1 Introducción.....	55
5.2 Arquitectura del sistema .....	55
5.3 Sistema base .....	55
5.3.1 Módulo Detector de Personas.....	56
5.3.1 Módulo <i>Tracker</i> .....	56
5.3.2 Módulo de <i>matching</i> básico.....	56
5.3.2.1 Inicialización de objetivos .....	56
5.3.2.2 Asociación de objetivos.....	57
5.3.2.3 Terminación de objetivos .....	60
5.3.2.4 Formato de salida y visualización del sistema .....	60
5.4 Ajuste de parámetros del sistema base a las características de los vídeos .....	61
5.4.1 Módulo Detector de Personas.....	61
5.4.1.1 Detector Latent SVM.....	61
5.4.2 Módulo Tracker .....	62
5.4.2.1 Filtro de Kalman.....	62
5.4.2.2 Filtro de Partículas basado en color.....	62
5.4.3 Módulo de <i>Matching</i> básico .....	62
5.5 Resultados obtenidos del sistema base .....	63
5.5.1 Resultados obtenidos utilizando el detector Latent SVM y el <i>tracker</i> Filtro de Kalman .....	63
5.5.2 Resultados obtenidos utilizando el detector latent SVM y el <i>tracker</i> Filtro de Partículas basado en color .....	64
5.5.3 Comparativa de resultados obtenidos .....	64
5.6 Conclusiones.....	65
6 Modificaciones del sistema base .....	67
6.1 Introducción.....	67
6.2 Modificaciones del módulo de <i>matching</i> básico del sistema base .....	67
6.2.1 Primera modificación del sistema base .....	67
6.2.1.1 Actualización del modelo del tracker utilizando únicamente la información del detector de personas.....	67
6.2.1.2 Resultados obtenidos para cada nivel del dataset .....	67
6.2.1.3 Comparativa de resultados.....	68
6.2.2 Segunda modificación del sistema base .....	70
6.2.2.1 Inicialización del tracker al detectarse que pierde al objetivo.....	70
6.2.2.2 Resultados obtenidos para cada nivel del dataset .....	71
6.2.2.3 Comparativa de resultados.....	71
6.2.3 Tercera modificación del sistema base .....	74

6.2.3.1 Inicialización directa de objetivos con detecciones de alta confianza.....	74
6.2.3.1 Resultados obtenidos para cada nivel del dataset .....	74
6.2.3.2 Comparativa de resultados.....	74
6.2.4 Cuarta modificación del sistema base .....	78
6.2.4.1 Búsqueda del parámetro óptimo $\beta$ .....	78
6.2.4.1 Resultados obtenidos para cada nivel del dataset .....	79
6.2.4.2 Comparativa de resultados.....	80
6.2.5 Quinta modificación del sistema base .....	83
6.2.5.1 Muerte directa de objetivos en las zonas de salida de la secuencia.....	83
6.2.5.2 Resultados obtenidos para cada nivel del dataset .....	84
6.2.5.3 Comparativa de resultados.....	84
6.2.6 Sexta modificación del sistema base .....	87
6.2.6.1 Búsqueda del parámetro óptimo $\alpha$ .....	87
6.2.6.2 Resultados obtenidos para cada nivel del dataset .....	88
6.2.6.3 Comparativa de resultados.....	89
6.3 Conclusiones.....	92
7 Generalización de resultados .....	95
7.1 Introducción.....	95
7.2 Resultados finales obtenidos .....	95
7.2.1 Resultados obtenidos utilizando el detector latent SVM y el <i>tracker</i> Filtro de Kalman .....	95
7.2.2 Resultados obtenidos utilizando el detector latent SVM y el <i>tracker</i> Filtro de Partículas basado en color .....	97
7.2.3 Comparativa de resultados finales .....	99
7.3 Conclusiones.....	101
8 Conclusiones y trabajo futuro.....	103
8.1 Conclusiones.....	103
8.2 Trabajo futuro .....	104
Referencias .....	105
Anexos .....	I
A. Resultados obtenidos del sistema base .....	I
A.1. Parámetros y resultados obtenidos utilizando el detector latent SVM y el <i>tracker</i> Filtro de Kalman para cada nivel del dataset.....	I
A.1.1 Nivel 1 .....	I
A.1.2 Nivel 2 .....	I
A.1.3 Nivel 3 .....	I
A.1.4 Nivel 4 .....	I
A.1.5 Nivel 5 .....	II
A.1.6 Nivel 6 .....	II
A.2. Parámetros y resultados obtenidos utilizando el detector latent SVM y el <i>tracker</i> Filtro de Partículas para cada nivel del dataset .....	II
A.2.1 Nivel 1 .....	II
A.2.2 Nivel 2 .....	II
A.2.3 Nivel 3 .....	II
A.2.4 Nivel 4 .....	III
A.2.5 Nivel 5 .....	III
A.2.6 Nivel 6 .....	III
B. Resultados obtenidos para cada una de las modificaciones del sistema baseV	
B.1. Primera modificación .....	V
B.1.1 Resultados obtenidos para cada nivel del dataset .....	V

B.1.1.1	Nivel 1 .....	V
B.1.1.2	Nivel 2 .....	V
B.1.1.3	Nivel 3 .....	V
B.1.1.4	Nivel 4 .....	V
B.1.1.5	Nivel 5 .....	V
B.1.1.6	Nivel 6 .....	V
B.2.	Segunda modificación .....	VI
B.2.1	Resultados obtenidos para cada nivel del dataset .....	VI
B.2.1.1	Nivel 1 .....	VI
B.2.1.2	Nivel 2 .....	VI
B.2.1.3	Nivel 3 .....	VI
B.2.1.4	Nivel 4 .....	VI
B.2.1.5	Nivel 5 .....	VI
B.2.1.6	Nivel 6 .....	VI
B.3.	Tercera modificación .....	VI
B.3.1	Resultados obtenidos para cada nivel del dataset .....	VI
B.3.1.1	Nivel 1 .....	VI
B.3.1.2	Nivel 2 .....	VI
B.3.1.3	Nivel 3 .....	VII
B.3.1.4	Nivel 4 .....	VII
B.3.1.5	Nivel 5 .....	VII
B.3.1.6	Nivel 6 .....	VII
B.4.	Cuarta modificación .....	VII
B.4.1	Resultados obtenidos para cada nivel del dataset en función del parámetro $\beta$ .....	VII
B.4.1.1	$\beta = 0.3$ .....	VII
B.4.1.2	$\beta = 0.4$ .....	VIII
B.4.1.3	$\beta = 0.5$ .....	IX
B.4.1.4	$\beta = 0.6$ .....	X
B.4.1.5	$\beta = 0.7$ .....	X
B.4.2	Comparativa de resultados medios obtenidos para cada nivel del dataset en función del parámetro $\beta$ .....	XI
B.5.	Quinta modificación .....	XII
B.5.1	Resultados obtenidos para cada nivel del dataset .....	XII
B.6.	Sexta modificación .....	XIII
B.6.1	Resultados obtenidos para cada nivel del dataset en función del parámetro $\alpha$ .....	XIII
B.6.1.1	$\alpha = 0.0$ .....	XIII
B.6.1.2	$\alpha = 0.1$ .....	XIII
B.6.1.3	$\alpha = 0.2$ .....	XIV
B.6.1.4	$\alpha = 0.3$ .....	XV
B.6.1.5	$\alpha = 0.4$ .....	XVI
B.6.1.6	$\alpha = 0.5$ .....	XVII
B.6.1.7	$\alpha = 0.6$ .....	XVII
B.6.1.8	$\alpha = 0.7$ .....	XVIII
B.6.1.9	$\alpha = 0.8$ .....	XIX
B.6.2	Comparativa de resultados medios obtenidos para cada nivel del dataset en función del parámetro $\alpha$ .....	XX
C.	Resultados finales obtenidos para la combinación del detector Latent SVM y el Filtro de Kalman .....	XXIII

C.1.	Nivel 1 .....	XXIII
C.2.	Nivel 2 .....	XXIII
C.3.	Nivel 3 .....	XXIII
C.4.	Nivel 4 .....	XXIII
C.5.	Nivel 5 .....	XXIII
C.6.	Nivel 6 .....	XXIII
D.	Refinamiento de los parámetros del Filtro de Partículas basado en color	XXV
D.1.	Asignación de $\sigma_x$ y $\sigma_y$ a un valor de 5 para las secuencias que pertenecen al <i>dataset</i> público PETS09 .....	XXV
D.1.1	Resultados obtenidos para cada una de las secuencias de los niveles del <i>dataset</i>	XXV
D.1.1.1	Nivel 1 .....	XXV
D.1.1.2	Nivel 2 .....	XXV
D.1.1.3	Nivel 3 .....	XXV
D.1.1.4	Nivel 4 .....	XXV
D.1.1.5	Nivel 5 .....	XXV
D.1.1.6	Nivel 6 .....	XXV
D.1.2	Resultados medios obtenidos para cada nivel del <i>dataset</i> .....	XXVI
D.2.	Asignación de $\sigma_x$ y $\sigma_y$ a un valor de 6 para las secuencias que pertenecen al <i>dataset</i> público PETS09 .....	XXVI
D.2.1	Resultados obtenidos para cada una de las secuencias de los niveles del <i>dataset</i>	XXVI
D.2.1.1	Nivel 1 .....	XXVI
D.2.1.2	Nivel 2 .....	XXVI
D.2.1.3	Nivel 3 .....	XXVI
D.2.1.4	Nivel 4 .....	XXVI
D.2.1.5	Nivel 5 .....	XXVI
D.2.1.6	Nivel 6 .....	XXVI
D.2.1	Resultados medios obtenidos para cada nivel del <i>dataset</i> .....	XXVII
D.3.	Asignación de $\sigma_x$ y $\sigma_y$ a un valor de 7 para las secuencias que pertenecen al <i>dataset</i> público PETS09 .....	XXVII
D.3.1	Resultados obtenidos para cada una de las secuencias de los niveles del <i>dataset</i>	XXVII
D.3.1.1	Nivel 1 .....	XXVII
D.3.1.2	Nivel 2 .....	XXVII
D.3.1.3	Nivel 3 .....	XXVII
D.3.1.4	Nivel 4 .....	XXVII
D.3.1.5	Nivel 5 .....	XXVII
D.3.1.6	Nivel 6 .....	XXVII
D.3.2	Resultados medios obtenidos para cada nivel del <i>dataset</i> .....	XXVIII
D.4.	Asignación de $\sigma_x$ y $\sigma_y$ a un valor de 8 para las secuencias que pertenecen al <i>dataset</i> público PETS09 .....	XXVIII
D.4.1	Resultados obtenidos para cada una de las secuencias de los niveles del <i>dataset</i>	XXVIII
D.4.1.1	Nivel 1 .....	XXVIII
D.4.1.1	Nivel 2 .....	XXVIII
D.4.1.1	Nivel 3 .....	XXVIII
D.4.1.1	Nivel 4 .....	XXVIII
D.4.1.1	Nivel 5 .....	XXVIII
D.4.1.1	Nivel 6 .....	XXVIII

D.4.2	Resultados medios obtenidos para cada nivel del dataset .....	XXIX
D.5.	Comparativa de resultados medios obtenidos en función de los valores de $\sigma_x$ y $\sigma_y$ .....	XXIX

## INDICE DE FIGURAS

FIGURA 2-1:	ARQUITECTURA CANÓNICA DEL DETECTOR DE PERSONAS (EXTRAÍDO DE [14]).....	4
FIGURA 2-2:	MÉTODOS DE DETECCIÓN DE OBJETOS (EXTRAÍDO DE [14]).....	5
FIGURA 2-3:	MÉTODOS DE CARACTERIZACIÓN DEL MODELO DE PERSONA (EXTRAÍDO DE [14]) .....	7
FIGURA 2-4:	REPRESENTACIÓN: A) CENTROIDE, B) MÚLTIPLES PUNTOS, C) <i>BOUNDING BOX</i> , D) ELIPSE, E) BASADO EN MÚLTIPLES PARTES, F) ESQUELETO DEL OBJETO, G) CONTORNO COMPLETO DEL OBJETO, H) PUNTOS DE CONTROL SOBRE EL CONTORNO DEL OBJETO, I) SILUETA DEL OBJETO. EXTRAÍDO DE [42].....	12
FIGURA 2-5:	MÉTODOS DE PREDICCIÓN DE MOVIMIENTO: A) BÚSQUEDA EXHAUSTIVA, B) BÚSQUEDA SELECTIVA (DONDE LOS PUNTOS AZULES REPRESENTAN LA SELECCIÓN).....	13
FIGURA 2-6:	CLASIFICACIÓN DE LOS ALGORITMOS DE SEGUIMIENTO DE OBJETOS (EXTRAÍDO DE [42]).....	14
FIGURA 2-7:	EJEMPLO DE LAS 8 PERSPECTIVAS DISPONIBLES EN PETS2009. ....	18
FIGURA 2-8:	EJEMPLO DE ALGUNOS <i>FRAMES</i> DE CAVIAR. ....	19
FIGURA 2-9:	EJEMPLO DE ALGUNOS <i>FRAMES</i> DE VISOR.....	20
FIGURA 2-10:	EJEMPLO DE ALGUNOS <i>FRAMES</i> DE I-LIDS. ....	21
FIGURA 3-1:	METODOLOGÍA DE EVALUACIÓN OBJETIVA PROPUESTA PARA EL SISTEMA DE SEGUIMIENTO DE MÚLTIPLES PERSONAS .....	29
FIGURA 3-2:	EJEMPLO DE ALGUNAS LÍNEAS DEL ARCHIVO GENERADO TRAS LA EJECUCIÓN DEL ALGORITMO DE SEGUIMIENTO .....	30
FIGURA 3-3:	EJEMPLO DE ALGUNAS LÍNEAS DE UN ARCHIVO DE ANOTACIÓN DEL <i>GROUND-TRUTH</i> . ....	31
FIGURA 3-4:	METODOLOGÍA DE EVALUACIÓN SUBJETIVA PROPUESTA PARA EL SISTEMA DE SEGUIMIENTO DE MÚLTIPLES PERSONAS. ....	31
FIGURA 3-5:	SALIDA VISUAL DEL <i>GROUND-TRUTH</i> Y DEL ALGORITMO DE SEGUIMIENTO DE UN OBJETIVO. ....	32
FIGURA 3-6:	EJEMPLO DE ALGUNOS <i>FRAMES</i> DE ALGUNA DE LAS SECUENCIAS DE LOS 3 PRIMEROS NIVELES DEL <i>DATASET</i> PROPUESTO (CADA FILA REPRESENTA UN NIVEL DEL <i>DATASET</i> EN ORDEN CRECIENTE DE DIFICULTAD). ....	35

FIGURA 3-7: EJEMPLO DE ALGUNOS <i>FRAMES</i> DE ALGUNA DE LAS SECUENCIAS DE LOS 3 ÚLTIMOS NIVELES DEL <i>DATASET</i> PROPUESTO (CADA FILA REPRESENTA UN NIVEL DEL <i>DATASET</i> EN ORDEN CRECIENTE DE DIFICULTAD). .....	36
FIGURA 4-1: DETECCIONES OBTENIDAS CON UN MODELO DE PERSONA DE UNA SOLA COMPONENTE. EL MODELO SE DEFINE MEDIANTE UN FILTRO RAÍZ (A), VARIOS FILTROS DE PARTES DE MAYOR RESOLUCIÓN (B) Y UN MODELO ESPACIAL PARA LA LOCALIZACIÓN DE CADA PARTE RELATIVA A LA RAÍZ (C). LOS FILTROS ESPECIFICAN PESOS PARA LAS CARACTERÍSTICAS DEL HISTOGRAMA DE GRADIENTES ORIENTADOS. SUS VISUALIZACIONES MUESTRAN LOS PESOS POSITIVOS EN DIFERENTES ORIENTACIONES. LA VISUALIZACIÓN DE LOS MODELOS ESPACIALES REFLEJA EL “COSTE” DE COLOCAR EL CENTRO DE UNA PARTE EN DIFERENTES LOCALIZACIONES RELATIVAS A LA RAÍZ. ....	40
FIGURA 4-2: UNA PIRÁMIDE DE CARACTERÍSTICAS Y UNA INSTANCIACIÓN DE UN MODELO DE PERSONA DENTRO DE DICHA PIRÁMIDE. LOS FILTROS DE PARTES SE COLOCAN AL DOBLE DE RESOLUCIÓN DE LA POSICIÓN DEL FILTRO RAÍZ. ....	41
FIGURA 4-3: EL PROCEDIMIENTO DE CÁLCULO DE LAS PUNTUACIONES GLOBALES EN UNA ESCALA. LAS RESPUESTAS DEL FILTRO RAÍZ Y DE LOS FILTROS DE PARTES SE CALCULAN A DIFERENTES RESOLUCIONES EN LA PIRÁMIDE DE CARACTERÍSTICAS. LAS RESPUESTAS TRANSFORMADAS SE COMBINAN PARA CONSEGUIR UNA PUNTUACIÓN FINAL PARA CADA COLOCACIÓN DE LA RAÍZ. SE MUESTRAN LAS RESPUESTAS Y LAS RESPUESTAS TRANSFORMADAS DE LAS PARTES DE LA “CABEZA” Y “HOMBRO DERECHO”. NÓTESE QUE LA “CABEZA” ES MÁS DISCRIMINATIVA. LAS RESPUESTAS COMBINADAS MUESTRAN DOS BUENAS HIPÓTESIS DEL OBJETO EN ESTA ESCALA. ....	43
FIGURA 4-4: PROCESO DEL FILTRO DE KALMAN (EXTRAÍDO DE [94]).....	47
FIGURA 4-5: LOS MÉTODOS DE LOCALIZACIÓN DE MÚLTIPLES HIPÓTESIS EXTRAEN UN NÚMERO VARIABLE DE HIPÓTESIS PARA DESPUÉS DETERMINARSE LA CALIDAD DE CADA MUESTRA (EXTRAÍDO DE [1]). ....	48
FIGURA 4-6: ESQUEMA GENERAL DEL PROCESO DE SEGUIMIENTO DE UN FILTRO DE PARTÍCULAS (EXTRAÍDO DE [95]). ....	51
FIGURA 5-1: DIAGRAMA DE BLOQUES DEL SISTEMA.....	55
FIGURA 5-2: ALGORITMO DE ASOCIACIÓN TIPO <i>GREEDY</i> . ....	60
FIGURA 5-3: EJEMPLO DE UNA IMAGEN DE LA SALIDA DEL DETECTOR DE PERSONAS LATENT SVM PARA UN UMBRAL DE CONFIANZA DE LA DETECCIÓN DE -0.598 (IZQUIERDA) Y 0.098 (DERECHA). ....	61
FIGURA 6-1: MOTP PARA DIFERENTES VALORES DEL PARÁMETRO B .....	78
FIGURA 6-2: MOTA PARA DIFERENTES VALORES DEL PARÁMETRO B.....	79
FIGURA 6-3: MOTP PARA DIFERENTES VALORES DEL PARÁMETRO A .....	87
FIGURA 6-4: MOTA PARA DIFERENTES VALORES DEL PARÁMETRO A.....	88

# INDICE DE TABLAS

TABLA 5-1: RESULTADOS MEDIOS OBTENIDOS PARA CADA NIVEL DEL <i>DATASET</i> UTILIZANDO EL DETECTOR LATENT SVM Y EL <i>TRACKER</i> FILTRO DE KALMAN. ....	64
TABLA 5-2: RESULTADOS MEDIOS OBTENIDOS PARA CADA NIVEL DEL <i>DATASET</i> UTILIZANDO EL DETECTOR LATENT SVM Y EL <i>TRACKER</i> FILTRO DE PARTÍCULAS.....	64
TABLA 5-3: COMPARATIVA DE RESULTADOS MEDIOS OBTENIDOS PARA CADA NIVEL DEL <i>DATASET</i> UTILIZANDO EL DETECTOR LATENT SVM Y LOS <i>TRACKERS</i> FILTRO DE KALMAN Y FILTRO DE PARTÍCULAS. LA PRIMERA COLUMNA DE CADA NIVEL MUESTRA LOS RESULTADOS OBTENIDOS UTILIZANDO EL <i>TRACKER</i> FILTRO DE KALMAN, LA SEGUNDA COLUMNA INDICA LOS RESULTADOS OBTENIDOS UTILIZANDO EL <i>TRACKER</i> FILTRO DE PARTÍCULAS Y LA TERCERA COLUMNA INDICA LA DIFERENCIA EN PORCENTAJE ENTRE LA MEJOR Y PEOR COMBINACIÓN DE DETECTOR- <i>TRACKER</i> EN CUANTO A LOS RESULTADOS OBTENIDOS PARA MOTA. ....	65
TABLA 6-1: RESULTADOS MEDIOS OBTENIDOS PARA CADA NIVEL DEL <i>DATASET</i> TRAS LA PRIMERA MEJORA INTRODUCIDA AL SISTEMA BASE. ....	68
TABLA 6-2: COMPARATIVA DE RESULTADOS MEDIOS OBTENIDOS EN LOS 3 PRIMEROS NIVELES DEL <i>DATASET</i> TRAS LA PRIMERA MEJORA INTRODUCIDA CON RESPECTO AL SISTEMA BASE. RESULTADOS ETAPA ANTERIOR SE CORRESPONDE CON LOS RESULTADOS OBTENIDOS PARA EL SISTEMA BASE, RESULTADOS TRAS MEJORA HACE REFERENCIA A LOS RESULTADOS OBTENIDOS TRAS LA MEJORA INTRODUCIDA, MEJORA INCREMENTAL TRAS LA MODIFICACIÓN ES EL PORCENTAJE DE MEJORA CON RESPECTO A LA MEJORA ANTERIOR (EN ESTE CASO, EL SISTEMA BASE), RESULTADOS SISTEMA BASE SON LOS RESULTADOS OBTENIDOS DEL SISTEMA BASE ORIGINAL Y MEJORA TOTAL SOBRE EL SISTEMA BASE ES EL PORCENTAJE DE MEJORA RESPECTO A LOS RESULTADOS OBTENIDOS DEL SISTEMA BASE ORIGINAL, ESTO ES, LA ACUMULACIÓN DE MEJORAS.....	69
TABLA 6-3: COMPARATIVA DE RESULTADOS MEDIOS OBTENIDOS EN LOS 3 ÚLTIMOS NIVELES DEL <i>DATASET</i> TRAS LA PRIMERA MEJORA INTRODUCIDA CON RESPECTO AL SISTEMA BASE. RESULTADOS ETAPA ANTERIOR SE CORRESPONDE CON LOS RESULTADOS OBTENIDOS PARA EL SISTEMA BASE, RESULTADOS TRAS MEJORA HACE REFERENCIA A LOS RESULTADOS OBTENIDOS TRAS LA MEJORA INTRODUCIDA, MEJORA INCREMENTAL TRAS LA MODIFICACIÓN ES EL PORCENTAJE DE MEJORA CON RESPECTO A LA MEJORA ANTERIOR (EN ESTE CASO, EL SISTEMA BASE), RESULTADOS SISTEMA BASE SON LOS RESULTADOS OBTENIDOS DEL SISTEMA BASE ORIGINAL Y MEJORA TOTAL SOBRE EL SISTEMA BASE ES EL PORCENTAJE DE MEJORA RESPECTO A LOS RESULTADOS OBTENIDOS DEL SISTEMA BASE ORIGINAL, ESTO ES, LA ACUMULACIÓN DE MEJORAS.....	70
TABLA 6-4: RESULTADOS MEDIOS OBTENIDOS PARA CADA NIVEL DEL <i>DATASET</i> TRAS LA SEGUNDA MEJORA INTRODUCIDA AL SISTEMA BASE. ....	71
TABLA 6-5: COMPARATIVA DE RESULTADOS MEDIOS OBTENIDOS EN LOS 3 PRIMEROS NIVELES DEL <i>DATASET</i> TRAS LA SEGUNDA MEJORA INTRODUCIDA CON RESPECTO AL SISTEMA BASE Y AL SISTEMA BASE MODIFICADO CON LAS MEJORAS ANTERIORES. RESULTADOS ETAPA ANTERIOR SE CORRESPONDE CON LOS RESULTADOS OBTENIDOS PARA EL SISTEMA BASE, RESULTADOS TRAS MEJORA HACE REFERENCIA A LOS RESULTADOS OBTENIDOS TRAS LA MEJORA INTRODUCIDA, MEJORA INCREMENTAL TRAS LA MODIFICACIÓN ES EL PORCENTAJE DE MEJORA CON RESPECTO A LA MEJORA ANTERIOR (EN ESTE CASO, EL SISTEMA BASE), RESULTADOS SISTEMA BASE SON LOS RESULTADOS OBTENIDOS DEL SISTEMA BASE ORIGINAL Y MEJORA TOTAL SOBRE EL SISTEMA BASE ES EL PORCENTAJE DE MEJORA RESPECTO A LOS	

RESULTADOS OBTENIDOS DEL SISTEMA BASE ORIGINAL, ESTO ES, LA ACUMULACIÓN DE MEJORAS. .... 72

TABLA 6-6: COMPARATIVA DE RESULTADOS MEDIOS OBTENIDOS EN LOS 3 ÚLTIMOS NIVELES DEL DATASET TRAS LA SEGUNDA MEJORA INTRODUCIDA CON RESPECTO AL SISTEMA BASE Y AL SISTEMA BASE MODIFICADO CON LAS MEJORAS ANTERIORES. RESULTADOS ETAPA ANTERIOR SE CORRESPONDE CON LOS RESULTADOS OBTENIDOS PARA EL SISTEMA BASE, RESULTADOS TRAS MEJORA HACE REFERENCIA A LOS RESULTADOS OBTENIDOS TRAS LA MEJORA INTRODUCIDA, MEJORA INCREMENTAL TRAS LA MODIFICACIÓN ES EL PORCENTAJE DE MEJORA CON RESPECTO A LA MEJORA ANTERIOR (EN ESTE CASO, EL SISTEMA BASE), RESULTADOS SISTEMA BASE SON LOS RESULTADOS OBTENIDOS DEL SISTEMA BASE ORIGINAL Y MEJORA TOTAL SOBRE EL SISTEMA BASE ES EL PORCENTAJE DE MEJORA RESPECTO A LOS RESULTADOS OBTENIDOS DEL SISTEMA BASE ORIGINAL, ESTO ES, LA ACUMULACIÓN DE MEJORAS. .... 73

TABLA 6-7: RESULTADOS MEDIOS OBTENIDOS PARA CADA NIVEL DEL DATASET TRAS LA TERCERA MEJORA INTRODUCIDA AL SISTEMA BASE. .... 74

TABLA 6-8: COMPARATIVA DE RESULTADOS MEDIOS OBTENIDOS EN LOS 3 PRIMEROS NIVELES DEL DATASET TRAS LA TERCERA MEJORA INTRODUCIDA CON RESPECTO AL SISTEMA BASE Y AL SISTEMA BASE MODIFICADO CON LAS MEJORAS ANTERIORES. RESULTADOS ETAPA ANTERIOR SE CORRESPONDE CON LOS RESULTADOS OBTENIDOS PARA EL SISTEMA BASE, RESULTADOS TRAS MEJORA HACE REFERENCIA A LOS RESULTADOS OBTENIDOS TRAS LA MEJORA INTRODUCIDA, MEJORA INCREMENTAL TRAS LA MODIFICACIÓN ES EL PORCENTAJE DE MEJORA CON RESPECTO A LA MEJORA ANTERIOR (EN ESTE CASO, EL SISTEMA BASE), RESULTADOS SISTEMA BASE SON LOS RESULTADOS OBTENIDOS DEL SISTEMA BASE ORIGINAL Y MEJORA TOTAL SOBRE EL SISTEMA BASE ES EL PORCENTAJE DE MEJORA RESPECTO A LOS RESULTADOS OBTENIDOS DEL SISTEMA BASE ORIGINAL, ESTO ES, LA ACUMULACIÓN DE MEJORAS. .... 76

TABLA 6-9: COMPARATIVA DE RESULTADOS MEDIOS OBTENIDOS EN LOS 3 ÚLTIMOS NIVELES DEL DATASET TRAS LA TERCERA MEJORA INTRODUCIDA CON RESPECTO AL SISTEMA BASE Y AL SISTEMA BASE MODIFICADO CON LAS MEJORAS ANTERIORES. RESULTADOS ETAPA ANTERIOR SE CORRESPONDE CON LOS RESULTADOS OBTENIDOS PARA EL SISTEMA BASE, RESULTADOS TRAS MEJORA HACE REFERENCIA A LOS RESULTADOS OBTENIDOS TRAS LA MEJORA INTRODUCIDA, MEJORA INCREMENTAL TRAS LA MODIFICACIÓN ES EL PORCENTAJE DE MEJORA CON RESPECTO A LA MEJORA ANTERIOR (EN ESTE CASO, EL SISTEMA BASE), RESULTADOS SISTEMA BASE SON LOS RESULTADOS OBTENIDOS DEL SISTEMA BASE ORIGINAL Y MEJORA TOTAL SOBRE EL SISTEMA BASE ES EL PORCENTAJE DE MEJORA RESPECTO A LOS RESULTADOS OBTENIDOS DEL SISTEMA BASE ORIGINAL, ESTO ES, LA ACUMULACIÓN DE MEJORAS. .... 77

TABLA 6-10: RESULTADOS MEDIOS OBTENIDOS PARA CADA NIVEL DEL DATASET TRAS LA CUARTA MEJORA INTRODUCIDA AL SISTEMA BASE. .... 79

TABLA 6-11: COMPARATIVA DE RESULTADOS MEDIOS OBTENIDOS EN LOS 3 PRIMEROS NIVELES DEL DATASET TRAS LA CUARTA MEJORA INTRODUCIDA CON RESPECTO AL SISTEMA BASE Y AL SISTEMA BASE MODIFICADO CON LAS MEJORAS ANTERIORES. RESULTADOS ETAPA ANTERIOR SE CORRESPONDE CON LOS RESULTADOS OBTENIDOS PARA EL SISTEMA BASE, RESULTADOS TRAS MEJORA HACE REFERENCIA A LOS RESULTADOS OBTENIDOS TRAS LA MEJORA INTRODUCIDA, MEJORA INCREMENTAL TRAS LA MODIFICACIÓN ES EL PORCENTAJE DE MEJORA CON RESPECTO A LA MEJORA ANTERIOR (EN ESTE CASO, EL SISTEMA BASE), RESULTADOS SISTEMA BASE SON LOS RESULTADOS OBTENIDOS DEL SISTEMA BASE ORIGINAL Y MEJORA TOTAL SOBRE EL SISTEMA BASE ES EL PORCENTAJE DE MEJORA RESPECTO A LOS

RESULTADOS OBTENIDOS DEL SISTEMA BASE ORIGINAL, ESTO ES, LA ACUMULACIÓN DE MEJORAS. ....	81
TABLA 6-12: COMPARATIVA DE RESULTADOS MEDIOS OBTENIDOS EN LOS 3 ÚLTIMOS NIVELES DEL DATASET TRAS LA CUARTA MEJORA INTRODUCIDA CON RESPECTO AL SISTEMA BASE Y AL SISTEMA BASE MODIFICADO CON LAS MEJORAS ANTERIORES. RESULTADOS ETAPA ANTERIOR SE CORRESPONDE CON LOS RESULTADOS OBTENIDOS PARA EL SISTEMA BASE, RESULTADOS TRAS MEJORA HACE REFERENCIA A LOS RESULTADOS OBTENIDOS TRAS LA MEJORA INTRODUCIDA, MEJORA INCREMENTAL TRAS LA MODIFICACIÓN ES EL PORCENTAJE DE MEJORA CON RESPECTO A LA MEJORA ANTERIOR (EN ESTE CASO, EL SISTEMA BASE), RESULTADOS SISTEMA BASE SON LOS RESULTADOS OBTENIDOS DEL SISTEMA BASE ORIGINAL Y MEJORA TOTAL SOBRE EL SISTEMA BASE ES EL PORCENTAJE DE MEJORA RESPECTO A LOS RESULTADOS OBTENIDOS DEL SISTEMA BASE ORIGINAL, ESTO ES, LA ACUMULACIÓN DE MEJORAS. ....	82
TABLA 6-13: DEFINICIÓN DE LAS ZONAS DE SALIDA EN CADA UNA DE LAS SECUENCIAS DEL NIVEL .....	84
TABLA 6-14: RESULTADOS MEDIOS OBTENIDOS PARA CADA NIVEL DEL DATASET TRAS LA QUINTA MEJORA INTRODUCIDA AL SISTEMA BASE. ....	84
TABLA 6-15: COMPARATIVA DE RESULTADOS MEDIOS OBTENIDOS EN LOS 3 PRIMEROS NIVELES DEL DATASET TRAS LA QUINTA MEJORA INTRODUCIDA CON RESPECTO AL SISTEMA BASE Y AL SISTEMA BASE MODIFICADO CON LAS MEJORAS ANTERIORES. RESULTADOS ETAPA ANTERIOR SE CORRESPONDE CON LOS RESULTADOS OBTENIDOS PARA EL SISTEMA BASE, RESULTADOS TRAS MEJORA HACE REFERENCIA A LOS RESULTADOS OBTENIDOS TRAS LA MEJORA INTRODUCIDA, MEJORA INCREMENTAL TRAS LA MODIFICACIÓN ES EL PORCENTAJE DE MEJORA CON RESPECTO A LA MEJORA ANTERIOR (EN ESTE CASO, EL SISTEMA BASE), RESULTADOS SISTEMA BASE SON LOS RESULTADOS OBTENIDOS DEL SISTEMA BASE ORIGINAL Y MEJORA TOTAL SOBRE EL SISTEMA BASE ES EL PORCENTAJE DE MEJORA RESPECTO A LOS RESULTADOS OBTENIDOS DEL SISTEMA BASE ORIGINAL, ESTO ES, LA ACUMULACIÓN DE MEJORAS. ....	85
TABLA 6-16: COMPARATIVA DE RESULTADOS MEDIOS OBTENIDOS EN LOS 3 ÚLTIMOS NIVELES DEL DATASET TRAS LA QUINTA MEJORA INTRODUCIDA CON RESPECTO AL SISTEMA BASE Y AL SISTEMA BASE MODIFICADO CON LAS MEJORAS ANTERIORES. RESULTADOS ETAPA ANTERIOR SE CORRESPONDE CON LOS RESULTADOS OBTENIDOS PARA EL SISTEMA BASE, RESULTADOS TRAS MEJORA HACE REFERENCIA A LOS RESULTADOS OBTENIDOS TRAS LA MEJORA INTRODUCIDA, MEJORA INCREMENTAL TRAS LA MODIFICACIÓN ES EL PORCENTAJE DE MEJORA CON RESPECTO A LA MEJORA ANTERIOR (EN ESTE CASO, EL SISTEMA BASE), RESULTADOS SISTEMA BASE SON LOS RESULTADOS OBTENIDOS DEL SISTEMA BASE ORIGINAL Y MEJORA TOTAL SOBRE EL SISTEMA BASE ES EL PORCENTAJE DE MEJORA RESPECTO A LOS RESULTADOS OBTENIDOS DEL SISTEMA BASE ORIGINAL, ESTO ES, LA ACUMULACIÓN DE MEJORAS. ....	86
TABLA 6-17: RESULTADOS MEDIOS OBTENIDOS PARA CADA NIVEL DEL DATASET TRAS LA SEXTA MEJORA INTRODUCIDA AL SISTEMA BASE. ....	88
TABLA 6-18: COMPARATIVA DE RESULTADOS MEDIOS OBTENIDOS EN LOS 3 PRIMEROS NIVELES DEL DATASET TRAS LA SEXTA MEJORA INTRODUCIDA CON RESPECTO AL SISTEMA BASE Y AL SISTEMA BASE MODIFICADO CON LAS MEJORAS ANTERIORES. RESULTADOS ETAPA ANTERIOR SE CORRESPONDE CON LOS RESULTADOS OBTENIDOS PARA EL SISTEMA BASE, RESULTADOS TRAS MEJORA HACE REFERENCIA A LOS RESULTADOS OBTENIDOS TRAS LA MEJORA INTRODUCIDA, MEJORA INCREMENTAL TRAS LA MODIFICACIÓN ES EL PORCENTAJE DE	

MEJORA CON RESPECTO A LA MEJORA ANTERIOR (EN ESTE CASO, EL SISTEMA BASE), RESULTADOS SISTEMA BASE SON LOS RESULTADOS OBTENIDOS DEL SISTEMA BASE ORIGINAL Y MEJORA TOTAL SOBRE EL SISTEMA BASE ES EL PORCENTAJE DE MEJORA RESPECTO A LOS RESULTADOS OBTENIDOS DEL SISTEMA BASE ORIGINAL, ESTO ES, LA ACUMULACIÓN DE MEJORAS. .... 90

TABLA 6-19: COMPARATIVA DE RESULTADOS MEDIOS OBTENIDOS EN LOS 3 ÚLTIMOS NIVELES DEL DATASET TRAS LA SEXTA MEJORA INTRODUCIDA CON RESPECTO AL SISTEMA BASE Y AL SISTEMA BASE MODIFICADO CON LAS MEJORAS ANTERIORES. RESULTADOS ETAPA ANTERIOR SE CORRESPONDE CON LOS RESULTADOS OBTENIDOS PARA EL SISTEMA BASE, RESULTADOS TRAS MEJORA HACE REFERENCIA A LOS RESULTADOS OBTENIDOS TRAS LA MEJORA INTRODUCIDA, MEJORA INCREMENTAL TRAS LA MODIFICACIÓN ES EL PORCENTAJE DE MEJORA CON RESPECTO A LA MEJORA ANTERIOR (EN ESTE CASO, EL SISTEMA BASE), RESULTADOS SISTEMA BASE SON LOS RESULTADOS OBTENIDOS DEL SISTEMA BASE ORIGINAL Y MEJORA TOTAL SOBRE EL SISTEMA BASE ES EL PORCENTAJE DE MEJORA RESPECTO A LOS RESULTADOS OBTENIDOS DEL SISTEMA BASE ORIGINAL, ESTO ES, LA ACUMULACIÓN DE MEJORAS. .... 91

TABLA 7-1: RESULTADOS MEDIOS OBTENIDOS PARA CADA NIVEL DEL DATASET UTILIZÁNDOSE EL TRACKER FILTRO DE KALMAN TRAS LAS MEJORAS INTRODUCIDAS AL SISTEMA BASE. .... 96

TABLA 7-2: COMPARATIVA DE RESULTADOS MEDIOS OBTENIDOS PARA CADA UNO DE LOS NIVELES DEL DATASET ENTRE EL SISTEMA BASE Y EL SISTEMA FINAL CON LAS MEJORAS INTRODUCIDAS UTILIZÁNDOSE EL DETECTOR LATENT SVM Y EL FILTRO DE KALMAN. RESULTADOS SISTEMA BASE SON LOS RESULTADOS OBTENIDOS DEL SISTEMA BASE ORIGINAL, RESULTADOS TRAS MEJORAS HACE REFERENCIA A LOS RESULTADOS FINALES OBTENIDOS TRAS LAS MEJORAS INTRODUCIDAS Y MEJORA TOTAL SOBRE EL SISTEMA BASE ES EL PORCENTAJE DE MEJORA EN LOS RESULTADOS FINALES CON RESPECTO A LOS RESULTADOS OBTENIDOS DEL SISTEMA BASE ORIGINAL, ESTO ES, LA ACUMULACIÓN DE MEJORAS. .... 97

TABLA 7-3: RESULTADOS MEDIOS OBTENIDOS PARA CADA NIVEL DEL DATASET UTILIZÁNDOSE EL TRACKER FILTRO DE PARTÍCULAS TRAS LAS MEJORAS INTRODUCIDAS AL SISTEMA BASE..... 98

TABLA 7-4: COMPARATIVA DE RESULTADOS MEDIOS OBTENIDOS PARA CADA UNO DE LOS NIVELES DEL DATASET ENTRE EL SISTEMA BASE Y EL SISTEMA FINAL CON LAS MEJORAS INTRODUCIDAS UTILIZÁNDOSE EL DETECTOR LATENT SVM Y EL FILTRO DE PARTÍCULAS. RESULTADOS SISTEMA BASE SON LOS RESULTADOS OBTENIDOS DEL SISTEMA BASE ORIGINAL, RESULTADOS TRAS MEJORAS HACE REFERENCIA A LOS RESULTADOS FINALES OBTENIDOS TRAS LAS MEJORAS INTRODUCIDAS Y MEJORA TOTAL SOBRE EL SISTEMA BASE ES EL PORCENTAJE DE MEJORA EN LOS RESULTADOS FINALES CON RESPECTO A LOS RESULTADOS OBTENIDOS DEL SISTEMA BASE ORIGINAL, ESTO ES, LA ACUMULACIÓN DE MEJORAS. .... 99

TABLA 7-5: COMPARATIVA DE RESULTADOS MEDIOS OBTENIDOS PARA CADA NIVEL DEL DATASET UTILIZANDO EL DETECTOR LATENT SVM Y LOS TRACKERS FILTRO DE KALMAN Y FILTRO DE PARTÍCULAS CON LAS MEJORAS INTRODUCIDAS. LA PRIMERA COLUMNA DE CADA NIVEL MUESTRA LOS RESULTADOS OBTENIDOS UTILIZANDO EL TRACKER FILTRO DE KALMAN, LA SEGUNDA COLUMNA INDICA LOS RESULTADOS OBTENIDOS UTILIZANDO EL TRACKER FILTRO DE PARTÍCULAS Y LA TERCERA COLUMNA INDICA LA DIFERENCIA EN PORCENTAJE ENTRE LA MEJOR Y PEOR COMBINACIÓN DE DETECTOR-TRACKER EN CUANTO A LOS RESULTADOS OBTENIDOS PARA LA MEDIDA DE RENDIMIENTO MOTA. .... 101



# 1 Introducción

---

## 1.1 Motivación

El seguimiento de personas en secuencias de vídeo es un tema de análisis de gran interés en numerosas tareas de visión por computador como interfaces hombre-máquina, video-vigilancia, o en reconocimiento de acciones. Debido al gran interés mostrado por parte de la comunidad investigadora, actualmente se dispone de una gran cantidad de algoritmos de seguimiento (*trackers*) [1][2]. La mayoría de los algoritmos de seguimiento de múltiples personas utilizan el seguimiento basado en detección [3][4], que resuelve el problema en dos etapas:

- La detección de los objetos de interés (aquéllos en movimiento que no pertenecen al fondo de la escena) que indica las regiones de la imagen a seguir.
- La asociación de identidades consistentes con diferentes instancias del mismo objeto en el tiempo.

En el caso de personas, para la primera etapa se ha propuesto recientemente el uso de detectores específicos de personas cuyos resultados son utilizados como las regiones para llevar a cabo el seguimiento de personas debido a que la utilización de esta técnica ha demostrado una gran efectividad en escenas del mundo real complejas y gran densidad de personas. En [5][6][7] se presenta un estudio exhaustivo del estado del arte de detección de personas en diferentes aplicaciones de vídeo-vigilancia.

Por otro lado, en la segunda etapa, se pueden identificar diferentes enfoques clásicos que permiten la asociación de identidades en el tiempo como el *Joint Probabilistic Data Association Filter* (JPDAF) [8], el *Multi Hypotheses Tracking* (MHT) [9], el algoritmo Húngaro [10] o un algoritmo de tipo *greedy* [11].

Por tanto, la motivación de este proyecto surge por el continuo avance producido en análisis de vídeo que ha permitido el desarrollo de aplicaciones más ricas en diversos ámbitos como el de la vídeo-vigilancia. Además, la rápida mejoría tanto en la calidad y la resolución de los sensores de imagen, y el aumento espectacular en la potencia de cálculo de la última década han favorecido la creación de nuevos algoritmos y aplicaciones que utilizan el seguimiento de vídeo.

## 1.2 Objetivos

El objetivo de este proyecto es el diseño e implementación de un sistema de seguimiento de múltiples personas basado en la combinación de la información proporcionada por un detector de personas y un *tracker*. Se trata, en primer lugar, de tener un sistema base sobre el que aplicar mejoras en su módulo de asociación de identidades. Para ello, se consideran los siguientes sub-objetivos:

- Estudio detallado del estado del arte de detección de personas, seguimiento de objetos en vídeo y seguimiento de múltiples personas basado en detección.

- Selección de algoritmos relevantes de detección de personas y de seguimiento (*trackers*).
- Generar un algoritmo base de seguimiento de múltiples personas que utilice un proceso de *matching* para realizar la asociación de detecciones entre *frames* consecutivos.
- Evaluación del sistema implementado utilizando métricas objetivas, haciendo uso de *datasets* con *ground-truth* y de las métricas seleccionadas.
- Finalmente, una vez implementado el sistema básico, añadir mejoras al proceso de *matching* del sistema base y su posterior evaluación frente a los resultados obtenidos del primer sistema.

### 1.3 Organización de la memoria

La memoria del proyecto consta de los siguientes capítulos:

- **Capítulo 1:** Introducción y motivación del proyecto.
- **Capítulo 2:** Estudio detallado del estado del arte del seguimiento de múltiples personas en vídeo basado en detección.
- **Capítulo 3:** Explica el protocolo de evaluación propuesto para evaluar el algoritmo de seguimiento múltiple a desarrollar.
- **Capítulo 4:** Describe en profundidad los algoritmos que han sido seleccionados para llevar a cabo el seguimiento de múltiples personas en vídeo en este proyecto.
- **Capítulo 5:** Presenta el sistema base que se ha desarrollado para el seguimiento de múltiples personas basado en detección, los ajustes realizados y los resultados obtenidos.
- **Capítulo 6:** Incluye una explicación de las mejoras introducidas al sistema base desarrollado inicialmente, los resultados obtenidos tras cada una de las modificaciones y una comparativa para medir la mejora introducida con respecto al sistema anterior.
- **Capítulo 7:** Incluye y describe los resultados finales obtenidos una vez añadidas cada una de las mejoras introducidas al sistema base para cada una de las combinaciones entre detector y *tracker* utilizadas en el proyecto.
- **Capítulo 8:** Conclusiones y trabajo futuro.
- **Referencias y anexos.**

## **2 Estado del arte**

---

### **2.1 Introducción**

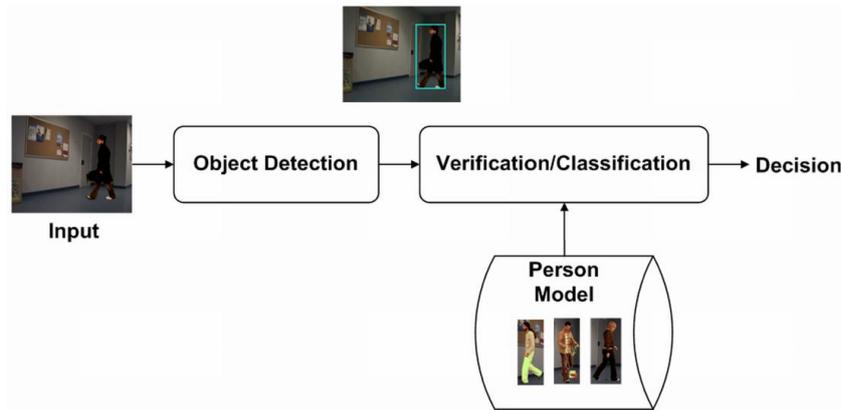
En este capítulo se presenta un estudio detallado del estado del arte del seguimiento de múltiples personas en vídeo basado en detección. Para ello, en primer lugar, se han estudiado los algoritmos de detección de personas existentes en la literatura actual (sección 2.2). Seguidamente, se han analizado, también, los algoritmos de seguimiento de objetos presentes en el estado del arte y sus características (sección 2.3). Posteriormente, se han resumido los algoritmos de seguimiento de múltiples personas existentes que combinan el seguimiento con la detección (sección 2.4), así como, los *datasets* públicos y métricas presentes en la literatura que permiten la evaluación de dichos algoritmos (sección 2.5). Por último, se han extraído una serie de conclusiones de cada una de las secciones que contiene el capítulo (sección 2.6).

### **2.2 Detección de personas**

La detección automática de personas en secuencias de vídeo [5][6][7] es uno de los problemas más desafiantes en visión por computador. La complejidad del problema de detección de personas se basa principalmente en la dificultad del modelado de las personas debido a su gran variabilidad en la apariencia física, partes del cuerpo articuladas, poses, movimientos, puntos de vista y las interacciones entre las diferentes personas y objetos. Esta complejidad es aún mayor en escenarios de vigilancia típicos del mundo real tales como aeropuertos, centros comerciales, etc., que a menudo incluyen una pluralidad de personas, múltiples oclusiones y variabilidad del fondo.

#### **2.2.1 Arquitectura de los sistemas de detección de personas**

Como definieron [12][13] para los sistemas canónicos de vigilancia, todos los enfoques de detección de personas consiste sobre todo, en primer lugar, del diseño y la formación (si el entrenamiento lo requiere) de un modelo de persona en base a unos parámetros característicos (movimiento, dimensiones, silueta, etc.) y, en segundo lugar, del ajuste de este modelo de persona a los candidatos a ser persona en la escena. Todos los candidatos que se ajustan al modelo serán detectados o clasificados como persona, mientras que todos los demás no serán detectados ni clasificados como persona. La Figura 2-1 muestra la arquitectura básica de cualquier detector de personas.



**Figura 2-1: Arquitectura canónica del detector de personas (extraído de [14])**

A continuación, se describen cada una de las etapas de la arquitectura básica de un detector de personas:

- **Entrada:** Hay muchos formatos posibles de entrada, que determinan el tipo de información de entrada disponible para el detector. En relación a la visión por computador, la unidad de entrada de procesamiento básico es la imagen o el *frame* en el caso de procesamiento de vídeo. Las imágenes de entrada pueden ser de varias resoluciones, 2D o 3D, color o escala de grises, el espectro visible o en el infrarrojo, etc. Los vídeos de entrada pueden ser de cámaras estáticas o móviles, mono o estéreo-visión, etc.
- **Detección de objetos:** La detección de objetos consiste en la generación o extracción de las hipótesis iniciales de objetos de la escena, es decir, los candidatos a ser una persona. Esta es una tarea crítica para la detección de personas. El enfoque elegido (por ejemplo, la sustracción de fondo, ventana deslizante) será muy determinante para algunos factores de rendimiento global en el proceso de detección: la velocidad de procesamiento, los resultados de detección, de robustez a variaciones en la escena, etc.
- **Modelo de persona:** El modelo de persona define las características y normas que los objetos deben cumplir en la escena con el fin de ser considerados como personas. Al igual que en el paso anterior, esto también es una tarea crítica para la detección de personas. El enfoque elegido (por ejemplo, holístico, basado en partes) será muy determinante en algunos factores globales de rendimiento en el proceso de detección: la velocidad de procesamiento, la robustez a variaciones de pose, oclusiones parciales, etc.
- **Verificación o clasificación:** La tarea de clasificación o verificación puede ser considerada como un reconocimiento estándar de los patrones en cuestión. Este proceso compara previamente modelos de objetos previamente entrenados y el modelo de objeto generado de una imagen o secuencia.
- **Decisión:** De acuerdo a la comparación o similitud calculada en la etapa anterior, se debe tomar una decisión final. Dependiendo de la posterior aplicación, la decisión puede ser binaria (persona o no persona) o difusa (un valor de confianza o probabilidad de ser una persona).

## 2.2.2 Clasificación de los algoritmos de detección de personas

Esta sección describe la clasificación de los algoritmos de personas realizada y describe los diferentes algoritmos representativos del estado del arte. Muchos criterios se pueden utilizar para clasificar los algoritmos de detección de personas; por ejemplo, las técnicas utilizadas (por ejemplo, la extracción de fondo o primer plano, estimación o compensación de movimiento), el tipo de modelos que se utilizan (por ejemplo, estadísticos, movimiento), el uso de la información en 2D o 3D, la modalidad del sensor (por ejemplo, luz visible, infrarrojo), la multiplicidad de sensores (monocular, estéreo o multicámara), la colocación del sensor (centralizado frente a distribuido), la movilidad del sensor (estacionario frente a movimiento), etc.

Como se ha mencionado en la sección anterior, las dos tareas principales y críticas de la detección de personas (detección de objetos y modelo de persona) determinan el rendimiento global de la detección; por ello, se ha decidido realizar una clasificación de los algoritmos del estado del arte de acuerdo a estas tareas. Primeramente, se clasifican los algoritmos de detección de personas de acuerdo al método utilizado para generar o extraer los objetos iniciales que son candidatos a ser una persona y seguidamente, se clasifican en función del modelo de persona elegido.

### 2.2.2.1 Métodos de detección de objetos

Hay dos métodos principales de detección de objetos (ver Figura 2-2): los que se basan en algún tipo de segmentación de la escena en objetos de primer plano y fondo y los que se basan en un escaneo exhaustivo. Hay algunos métodos que intentan combinar ambos métodos juntos. En cualquier caso, el resultado de esta etapa es la localización y dimensión (*bounding box*) de los diferentes objetos candidatos a ser persona de la escena.

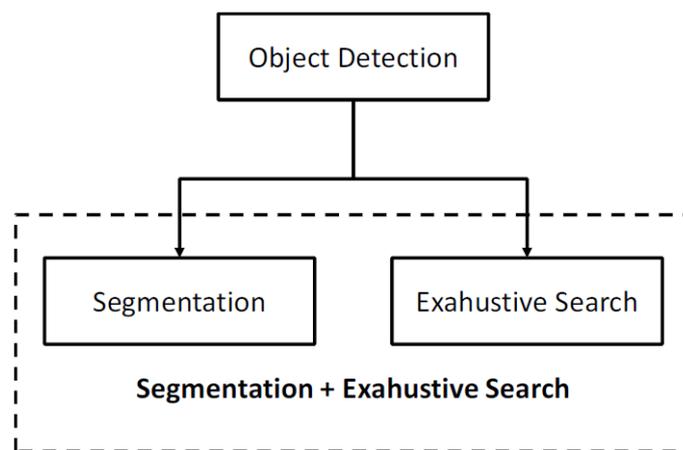


Figura 2-2: Métodos de detección de objetos (extraído de [14])

#### 2.2.2.1.1 Segmentación

La segmentación de imagen se utiliza a menudo para dividir una imagen en regiones separadas, que idealmente corresponden a diferentes objetos del mundo real. Más precisamente, es el proceso de asignar una etiqueta a cada píxel de una imagen de tal

manera que los píxeles con la etiqueta comparten alguna característica visual determinada o propiedad, como el color, el movimiento, la intensidad, textura, etc. Regiones adyacentes deben ser significativamente diferentes con respecto a la misma característica. El resultado final e ideal es localizar y discriminar objetos de la escena (primer plano) frente al resto de la imagen (fondo).

Actualmente, existen muchos enfoques del estado del arte que utilizan algún tipo de segmentación como un paso preliminar en la tarea de detección de personas. En particular, el uso de sustracción de fondo es muy popular en las aplicaciones de vigilancia [15][16][17][18][19][20][21]. Ellos tratan de detectar objetos en movimiento a partir de la diferencia entre el *frame* actual y un *frame* de referencia (modelo de fondo) y umbralizan los resultados para generar los objetos de interés. Hay algunos métodos que utilizan la segmentación de color [22][23], debido al hecho de que el color de la piel facilita la segmentación de la gente y el proceso de detección. Existen varios enfoques que utilizan algún tipo de información 3D para facilitar la segmentación por estéreo-visión [24][25][26] o directamente con cámaras 3D [27][28].

En relación a la detección de personas, el uso de la segmentación genera directamente los objetos candidatos a ser una persona y rechaza las áreas fácilmente irrelevantes de la imagen, es decir, sin los objetos de interés. Por esta razón, la tarea de clasificación posterior está claramente simplificada y, por lo tanto, el modelo de persona por lo general es más sencillo y tiene menor coste computacional. Sin embargo, ya que hay una fuerte dependencia con la segmentación, todos los problemas de segmentación se heredan (bajo y sobre segmentación). Estos problemas pueden afectar el rendimiento global de detección, principalmente limitando la tasa de detección máxima (objetos no detectados), pero también aumentando el número de falsas detecciones (detecciones de objetos parciales u objetos superpuestos). Además, estos problemas se magnifican en escenarios complejos donde es bastante difícil obtener una segmentación fiable.

### **2.2.2.1.2 Búsqueda exhaustiva**

La otra técnica para obtener hipótesis iniciales de localización de objetos es la búsqueda exhaustiva. Por lo general, consiste en la exploración de la imagen completa en busca de similitudes con el modelo de persona elegido en múltiples escalas y ubicaciones. A través de este mecanismo se obtiene un mapa de confianza denso o volumen (escala y localización). Con el fin de llegar a detecciones individuales, estos métodos deben buscar máximos locales en el volumen de densidad y, a continuación, aplicar algún tipo de supresión no máxima. Hay muchos enfoques de detección de personas en el estado del arte que utilizan esta técnica, de hecho, esta técnica es actualmente la más utilizada. Dentro de esta técnica, dos métodos diferentes se pueden utilizar. Por un lado, hay algunos enfoques que obtienen este volumen de densidad muestreando de forma implícita en un cuadrícula discreta 3D (localización y escala) mediante la evaluación de diferentes ventanas de detección con un clasificador. Este es el caso de la utilización de detectores basados en ventana deslizante, tales como [29][30][31][32][34][33]. Por otro lado, existen algunos enfoques que crean explícitamente este volumen de densidad a través de votos probabilísticos emitidos por la coincidencia de características locales. Este es el caso de los detectores basados en características tales como [3][35][36][37][38][39].

Generalmente, los detectores que utilizan este tipo de enfoques son más robustos a cambios de escala y variaciones de pose y, por lo tanto, más fiables en entornos complejos que los basados en la segmentación. Sin embargo, a diferencia de los otros enfoques, la tarea de clasificación no se simplifica, es aún más compleja debido a que el modelo de persona debe ser capaz de clasificar correctamente a un gran número de ejemplos negativos (posibles detecciones de falsos positivos). Además del aumento de la complejidad del modelo persona, el proceso de búsqueda exhaustiva en sí, por lo general, requiere un coste computacional superior, lo que hace que difícil cumplir requisitos de tiempo real.

### 2.2.2.1.3 Segmentación y búsqueda exhaustiva

Otro enfoque es la combinación de ambas técnicas de tal forma que se intente aprovechar sus fortalezas frente a sus inconvenientes. En [25], una selección inicial de candidatos se realiza utilizando la segmentación con información 3D y, a continuación, una segunda selección se realiza mediante búsqueda exhaustiva, pero debido a la eficiencia computacional sólo alrededor del centro de los candidatos pre-seleccionados, es decir, limitando la ventana deslizable.

### 2.2.2.1 Modelo de persona

Como ya hemos comentado, el proceso de verificación o clasificación aplica un modelo de persona previamente definido o entrenado a los objetos candidatos a ser una persona de una imagen o secuencia y toma una decisión final en función de su similitud (véase la Figura 2-1). Por lo tanto, la definición de un modelo de persona adecuado es una tarea fundamental para el proceso de verificación o clasificación. Hay dos principales fuentes de información discriminativas para caracterizar el modelo de las personas: la apariencia y el movimiento (ver Figura 2-3). En cualquier caso, el modelo debe ser capaz de discriminar entre las personas y cualquier otro objeto en la escena.

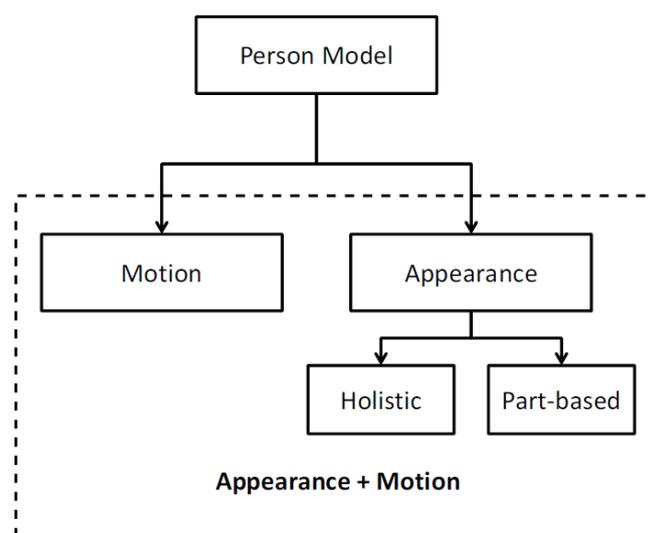


Figura 2-3: Métodos de caracterización del modelo de persona (extraído de [14])

### **2.2.2.1.1 Basado en movimiento**

Hoy en día en la literatura existente, la mayoría de los métodos se basan sólo en la información de apariencia o agregan solidez a la detección con la información de movimiento a través de algoritmos de seguimiento. Sin embargo, la apariencia humana varía debido a factores ambientales, tales como las condiciones de luz, ropa, contraste, etc., aparte de la enorme variabilidad intrínseca de las personas tales como diferentes alturas, anchuras, poses, etc. Por estas razones, hay algunos enfoques que tratan de evitar estos factores y realizan la detección utilizando sólo la información de movimiento [31][40].

Dentro de esta clasificación, [31] propone un sistema de clasificación de objetos basado en el análisis de movimiento periódico. El algoritmo segmenta el movimiento, sigue los objetos del primer plano, alinea cada objeto a lo largo del tiempo y finalmente calcula la auto-similitud entre objetos y cómo evoluciona en el tiempo. Otro enfoque basado en la información de movimiento [40] propone un sistema de detección de personas sobre la base de la detección de patrones de movimiento de las personas.

En relación a la detección de personas, los métodos basados en movimiento suelen obtener peores resultados que los métodos basados en apariencia, pero son independientes de la variabilidad de la apariencia. No soportan oclusiones parciales porque en este caso no se pueden extraer los patrones de movimiento correctamente. Por estas razones, sólo pueden ser considerados como información complementaria o en escenarios donde los métodos basados en apariencia no funcionan (por ejemplo, mala iluminación, pequeños objetos, etc.).

### **2.2.2.1.2 Basado en apariencia**

Hay muchos enfoques que utilizan la información de apariencia para definir el modelo de persona. Esto se debe a que la apariencia es más discriminante que el movimiento. Se han clasificado los modelos de persona de acuerdo a modelos humanos simplificados o modelos complejos. Existen modelos simples de personas que define a la persona como una región o forma, es decir, los modelos holísticos como [17][18][19][24][27][29][30][32][34][35][36] y modelos más complejos que definen a la persona como combinación de varias regiones o formas, es decir, los modelos basados en parte como [3][16][22][25][33][38].

Dentro de esta clasificación existen diferentes características escogidas para definir la apariencia de las personas, tanto en los modelos holísticos como en los modelos basados en parte. Hay algunos métodos que extraen la silueta objeto y clasifican el objeto de acuerdo con su similitud con las siluetas de personas de referencia o ciertas normas que la silueta que debe cumplir. Algunos enfoques hacen uso de la distribución del color en una persona (en los que el color de la piel es esencial) para determinar si el objeto es una persona o no. Pero los enfoques más populares son aquellos que definen la apariencia de las personas según su información de borde característica por medio de algún tipo de descriptor forma: características Haar-like, HOG (Histograma de Gradientes Orientados), *Edgelets* o ISM (Modelo de Forma Implícita).

En general, estos detectores basados en un modelo de persona simplificado o holístico tienen menor complejidad, pero no son compatibles con oclusiones parciales o variaciones

de pose. Si no se puede ver toda la región o la forma, el modelo no funciona correctamente. Por otro lado, los detectores basados en un modelo de persona más complejo o basado en partes, por lo general, tienen una mayor complejidad, pero soportan oclusiones parciales y variaciones de pose.

### ***2.2.2.1.3 Basado en movimiento y apariencia***

Aunque la gran mayoría de los enfoques se basan principalmente en la información de apariencia, hay algunos de los enfoques que combinan la apariencia y la información de movimiento con el fin de mejorar los resultados de la detección. Algunos autores combinan la apariencia y el movimiento expandiendo los detectores anteriores basados en apariencia a más de un *frame* como [29][41]. De esta manera son capaces de introducir fácilmente la información de movimiento en el modelo de persona y añaden robustez al detector. Últimamente, los enfoques más populares (enfoques de detección mediante seguimiento) son aquellos que combinan la detección y el seguimiento con el fin de mejorar los resultados de detección como [3][4][24][30][37]. En este caso, la información de movimiento no es implícitamente parte del modelo de persona, pero sigue siendo útil con el fin de filtrar o extrapolar las detecciones con el tiempo.

## **2.3 Seguimiento de objetos en vídeo**

El seguimiento de objetos en vídeo estudia el problema de la estimación de la trayectoria de un objeto de interés (objetivo) en una secuencia de imágenes. El seguimiento es una tarea complicada debido a la alta variabilidad y complejidad de los datos a analizar. En el diseño de un algoritmo de seguimiento, una serie de pasos se tienen en cuenta, tales como la extracción de características, la representación del objetivo (es decir, el modelado de objetos) y la propagación del modelo del objetivo en el tiempo. Con el fin de resolver los problemas mencionados anteriormente, se han desarrollado una gran variedad de técnicas que han sido resumidas en varios estudios publicados [1][42][43]. En esta sección, se ha sintetizado la información más relevante considerando las características, la representación, la predicción de movimiento y los algoritmos de seguimiento.

### **2.3.1 Características**

Todos los sistemas de seguimiento requieren definir las características que se usan para representar el objetivo. Es fundamental que las características elegidas identifiquen singularmente el objetivo del fondo de la escena u otros objetos, por lo tanto, ninguna característica se dice que es mejor que las otras, ya que depende de la situación en la que se utilizará el algoritmo de seguimiento. Como un ejemplo, si un jugador de fútbol en movimiento se desea seguir, el color no es la mejor característica ya que otros jugadores comparten esta característica. Por otra parte, la característica seleccionada tiene que ser descriptiva, así como flexible para hacer frente a cambios como rotaciones, la iluminación o cambios de escala. En la mayoría de los enfoques existentes, el usuario escoge manualmente estas características, en función de su experiencia y la aplicación final. Estas características se pueden agrupar en tres áreas: de bajo nivel, de nivel medio y de alto nivel [1].

### 2.3.1.1 Características de bajo nivel

A continuación, se describen las características de bajo nivel existentes:

**Color:** El color de una región está determinado por el valor medio de todos sus píxeles [42]. El problema con esta representación es que cuando se trata con distribuciones multimodales de color el valor medio no es suficiente. Por ejemplo, cuando se sigue una maleta roja el valor medio sería suficiente. Sin embargo, si una persona con esta maleta particular, tiene que ser seguida, la representación falla. Para resolver este problema, se introdujeron los histogramas de color (donde el número de veces que aparece un color se cuenta). Estos histogramas son ampliamente extendidos debido a su sencillez y buen rendimiento. Una de sus ventajas es que los cambios en la imagen, tales como rotación y traslación no les afectan, y si el objetivo gira, se ocluye o cambia su escala, el histograma no sufre un cambio significativo. Por otro lado, ya que los histogramas colapsan toda la información, es imposible saber qué color viene de qué parte de la imagen por lo que se pierde la estructura espacial del color del objeto. Además, distintos espacios de color existen (RGB, HSV, etc.), cada uno con sus ventajas y desventajas, por lo que no es posible elegir uno como el óptimo para todos los escenarios.

**Gradiente y derivados:** Los histogramas de gradientes orientados son una alternativa a los histogramas de color mencionados anteriormente, y muestran la distribución de la orientación de los vectores de gradiente dentro de la región [1]. Ellos son más robustos a los cambios de iluminación que los de color ya que los bordes tienden a persistir con dichos cambios. Sin embargo, el *clutter* afecta negativamente a la extracción del gradiente del objetivo.

### 2.3.1.2 Características de nivel medio

Mapear una imagen mediante características de bajo nivel puede no ser suficiente para lograr describir el contenido de dicha imagen, reduciendo así la eficacia de un *tracker* de vídeo. Una solución ampliamente utilizada consiste en analizar el vídeo usando subconjuntos de píxeles que representan estructuras relevantes (por ejemplo, los bordes y puntos de interés). Existen diversas técnicas que generan estos subconjuntos y después caracterizan dichas estructuras que pueden ser utilizadas para el seguimiento de vídeo.

**Bordes y líneas:** Bordes y líneas se han utilizado ampliamente a través de los años, y se han convertido en uno de los métodos más utilizados [44]. Un borde es un píxel situado en el límite entre dos regiones diferentes (cada uno con su intensidad constante). Los bordes están a menudo asociados con respuestas altas de los operadores de derivada de primer orden. Por ello, una forma muy simple de la caracterización de bordes consiste en la umbralización de la magnitud del gradiente de la imagen. Sin embargo, estas técnicas básicas presentan varios inconvenientes como la dificultad de la elección de un valor de umbral que permita sensibilidad a los bordes débiles y que a la vez evite falsas detecciones producidas por el ruido de la imagen o estructuras físicas fragmentadas (es decir, los píxeles etiquetados como bordes no están espacialmente conectados). Para explotar la continuidad espacial de los bordes, técnicas como el Detector de Canny [45] primero localizan un punto de borde inicial y luego propagan el resultado siguiendo la dirección normal del gradiente. Otra clase de métodos analiza variaciones de las derivadas de

segundo orden o de orden superior de la imagen. Una técnica popular de localización de borde [46] se basa en la detección de los pasos por cero del operador *LoG* (*Laplacian of Gaussian*) de la imagen que combina el Laplaciano con un Filtro Gaussiano. Para detectar cruces por cero horizontales y verticales, se puede umbralizar el valor absoluto de la convolución entre las filas y las columnas de la imagen filtrada *LoG* con un simple núcleo  $[-1 \ 1]$ .

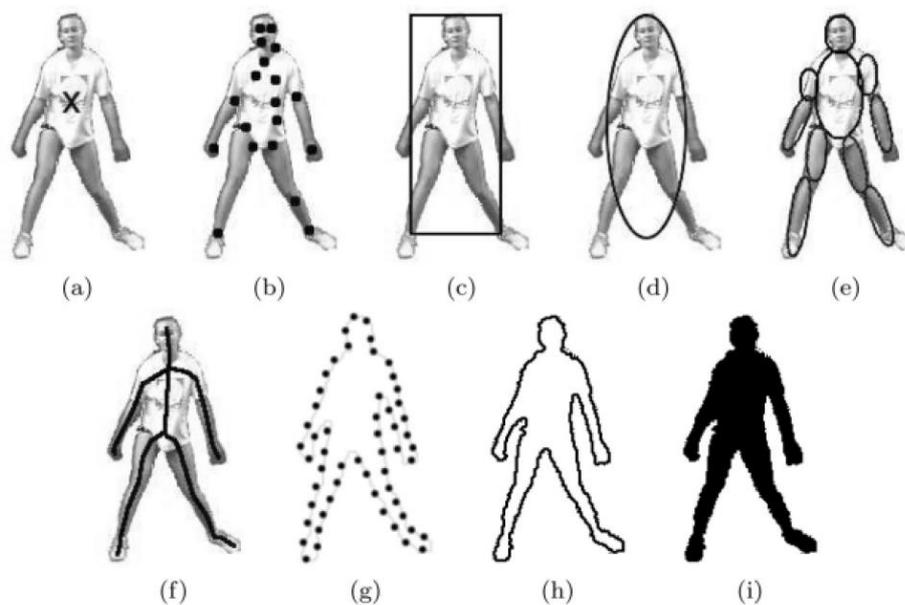
***Puntos de interés:*** Un punto de interés es un píxel en el que su estructura local de la imagen contiene información significativa para la imagen o el análisis de vídeo (por ejemplo, la esquina de una mesa, las articulaciones del cuerpo). Un ejemplo de las características utilizadas en el seguimiento son *Scale Invariant Feature Transform* (SIFT) [47] y *Speeded Up Robust Feature* (SURF) [48]. Por ejemplo, el detector SIFT es un detector de regiones de interés que extrae características a distintas escalas para lograr robustez a cambios de escalas del objetivo. Refina el procedimiento de localización por interpolación de la función *LoG* usando la expansión de Taylor de segundo orden. La misma aproximación se utiliza también para descartar regiones espurias cuya magnitud es menor que un umbral predefinido. Por último, picos pobremente definidos se eliminan por umbralización de la relación de los autovalores de la estructura del operador *LoG*.

### ***2.3.1.3 Características de alto nivel***

En lugar de la agrupación de características de nivel medio, otra opción para definir un objetivo es detectarlo entero sobre la base de su apariencia. Hay dos enfoques diferentes en este caso, dependiendo de la parte que se modela (fondo o primer plano) [1]. El primero determina qué parte de la imagen comprende el modelo fijado y el resto de objetos que no se puede explicar por medio de ese modelo se etiquetan como posibles objetivos. Entonces, estos objetos se siguen combinando características de bajo nivel [49]. En el segundo, la apariencia de una clase predefinida de objetivos se obtiene mediante el aprendizaje de características representativas de la clase seleccionada. Un ejemplo es la segmentación basada en color, utilizado para detectar las caras de las personas. Entonces, las caras detectadas se siguen de manera similar a como se describe en [50].

## **2.3.2 Representación**

Para proporcionar un seguimiento preciso en vídeo, las características del objetivo tienen que estar representadas adecuadamente por medio de un modelo. Para esta representación, hay varios enfoques: básico (donde el objetivo está representado por un solo punto), aproximaciones de área o volumen (donde las características se extraen de la zona seleccionada como, por ejemplo, los histogramas de color y plantillas de intensidad), articulados (donde una combinación de modelos rígidos aproximan la forma del objetivo), y deformable (donde se utilizan modelos fluidos, contornos o modelos de distribución punto). Algunos ejemplos de representación de forma se incluyen en la Figura 2-4. Por ejemplo, este modelo de objetivo puede contener información en cuanto a su forma y apariencia [1].



**Figura 2-4: Representación: a) centroide, b) múltiples puntos, c) *bounding box*, d) elipse, e) basado en múltiples partes, f) esqueleto del objeto, g) contorno completo del objeto, h) puntos de control sobre el contorno del objeto, i) silueta del objeto. Extraído de [42].**

### 2.3.3 Predicción de movimiento

La predicción de movimiento de un objetivo (también conocida como estimación de movimiento) determina el movimiento del objetivo realizando un análisis eficiente del seguimiento en el vídeo. Se puede considerar como una manera para reducir la carga computacional ya que adivinar donde el objetivo va a estar en cada *frame* ayuda a acelerar todo el proceso de seguimiento. La estimación de movimiento trata de explicar cómo el objetivo se mueve por medio de un vector de correspondencia (que describe el desplazamiento de un píxel entre dos *frames*) o por un campo de flujo óptico (que contiene información relativa a la velocidad si existe movimiento aparente).

Para localizar el objetivo en una determinada región de la imagen, se distinguen dos enfoques en función del área de búsqueda: búsqueda exhaustiva (determinista) o búsqueda selectiva (probabilística).

#### 2.3.3.1 Búsqueda exhaustiva

Este enfoque podría ser visto como un problema de optimización sin ningún conocimiento acerca del modelo de movimiento del objetivo. El objetivo es identificar y minimizar la función de coste que define la similitud entre el objetivo a ser seguido y las características observadas en el área de búsqueda del *frame* actual. Por lo tanto, se lleva a cabo una búsqueda exhaustiva para localizar el objetivo. La forma más elemental de usar esta información es comenzar el análisis en la misma posición que el target ocupaba en el *frame* anterior. Si la velocidad de *frames* de la secuencia es alta es obvia la ventaja de iniciar el análisis en la misma posición previamente determinada, ya que se disminuye significativamente el tiempo de búsqueda. Como se observa en la Figura 2-5(a), la caja azul representa el área de búsqueda. Se encuentra en la posición del objetivo en el *frame*

anterior (el objetivo se representa por medio de la caja más pequeña). Si el objetivo no cambia su posición bruscamente, este método proporciona buenos resultados y optimiza el tiempo de búsqueda. Este ejemplo corresponde a la aplicación del algoritmo MeanShift [51].

### 2.3.3.1 Búsqueda selectiva

Este enfoque supone un cierto modelo de movimiento del objetivo y consiste en dos etapas: predicción (teniendo en cuenta los datos anteriores) y actualización (teniendo en cuenta los datos de la imagen). Esta predicción utiliza el modelo de movimiento y proyecta hacia adelante el estado actual del objetivo (por ejemplo, la localización del objetivo y tamaño) del *frame* anterior al actual. La etapa de actualización comprueba las similitudes entre las predicciones y el modelo del objetivo (por ejemplo, la similitud del histograma de color) y no forma parte de la predicción del movimiento.

En la literatura actual, dos enfoques se utilizan ampliamente: Kalman [52] y Filtros de Partículas [53]. El Filtro de Kalman es una herramienta ideada para resolver problemas de estimación que son lineales y el ruido es Gaussiano. Los Filtros de Partículas permiten que el sistema (por ejemplo, el objetivo) se comporte de forma no lineal sin suponer gaussianidad. En la Figura 2-5(b) el enfoque del Filtro de Partículas se representa como un ejemplo de la búsqueda selectiva. Las partículas (puntos azules) describen cada punto donde se realiza una nueva búsqueda.



**Figura 2-5: Métodos de predicción de movimiento: a) Búsqueda exhaustiva, b) Búsqueda selectiva (donde los puntos azules representan la selección).**

### 2.3.4 Clasificación de los algoritmos de seguimiento

Una combinación de diferentes clasificaciones se ha propuesto en [1], [42], y se representa en la Figura 2-6.

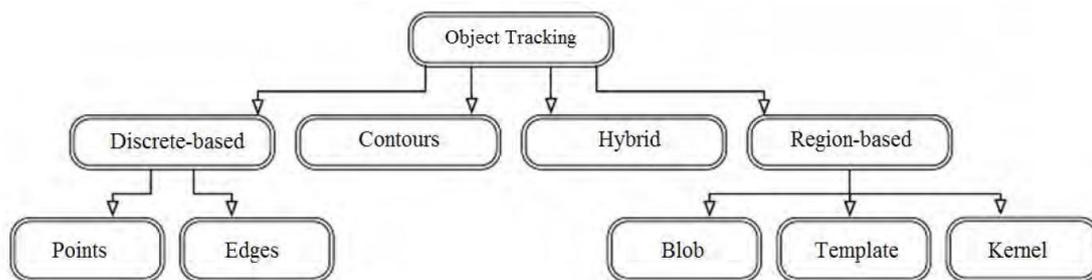


Figura 2-6: Clasificación de los algoritmos de seguimiento de objetos (extraído de [42]).

#### 2.3.4.1 Trackers basados en características discretas

Los *trackers* basados en características discretas utilizan estructuras de la imagen simples tales como puntos, líneas y bordes. Estos *trackers* utilizan puntos para representar los objetivos detectados en *frames* consecutivos y la asociación de los puntos se basa en el estado anterior del objetivo, que puede ser definido como la ubicación del objetivo y el movimiento [42].

**Puntos:** La información con respecto a la posición y el movimiento se incluye en el estado anterior. Para la detección del objetivo es necesario un mecanismo externo, y entonces los puntos característicos se extraen para representar el objetivo. Una de las mayores ventajas de los *trackers* de puntos es su capacidad para realizar un seguimiento de objetos pequeños (representados por un solo punto). Para cada punto, se calcula el coste de asociación entre el punto del *frame* anterior y el punto en el *frame* actual. Hay varias limitaciones que afectan a este coste: proximidad, velocidad, movimiento suave y rígido. Dado que este método proporciona soporte para varias situaciones diferentes ha sido ampliamente estudiado y hay varios trabajos en relación con el seguimiento de puntos deterministas [54].

**Bordes:** Se pueden distinguir dos enfoques diferentes: un modelo anterior se puede utilizar o no. Si no se utiliza un modelo, el seguimiento se realiza por medio de uno o más Filtros de Kalman y se repite en cada *frame*. Si se utiliza un modelo 3D del objeto es necesario desarrollar transformaciones complejas para alinear el modelo con el contenido de la imagen. Debido a su complejidad, los métodos de modelos 3D han visto reducida su utilización desde el comienzo de la década de 1990 [55].

#### 2.3.4.2 Trackers basados en contornos

Un contorno es una curva (abierta o cerrada) que describe un objetivo. La principal diferencia entre los *trackers* de contorno y los *trackers* de borde es que el último se centra en líneas rectas, mientras que el primero puede seguir objetivos con deformaciones de contorno. También hay que tener en cuenta que los enfoques de estas dos categorías son completamente diferentes, por lo que pertenecen a diferentes categorías. Los *trackers* de contorno se pueden dividir en tres categorías: básico (con aproximaciones de área o volumen), articulados (con modelos más complejos del objetivo incluyendo varios modelos rígidos) y deformables [1]. Se supone que hay sólo pequeños cambios en la forma y la posición del objetivo, así que la localización en cada *frame* se realiza cambiando ligeramente la información del *frame* anterior. Se han desarrollado varios enfoques durante

los años, incluyendo el uso de diferentes técnicas tales como conjuntos de nivel [56] y *trackers* que utilizan información de región [57].

#### 2.3.4.3 *Trackers híbridos*

Esta categoría representa una nueva tendencia que combina los enfoques descritos anteriormente con el objeto de mejorar el proceso de seguimiento global. En [58] una combinación de MeanShift y Filtro de Partículas se ha propuesto, donde el *tracker* produce un número menor de muestras usando Filtro de Partículas y luego desplaza las muestras hacia un máximo local utilizando MeanShift. Esto mejora la exactitud (con mejores resultados que la aplicación aislada de ambos algoritmos de seguimiento), mientras que se utilizan menos muestras que el Filtro de Partículas sólo.

#### 2.3.4.4 *Trackers basados en regiones*

Los *trackers* basados en región son conocidos por la gran variedad de características que utilizan para representar y realizar el seguimiento del objetivo. Algunas de estas características son el color, la textura, la intensidad y el gradiente. Los *trackers* pueden dividirse en tres grupos: *blob*, *template*, y *kernel*.

***Blob:*** Los *trackers* de *Blob* utilizan información muy específica y detalles del objetivo seguido, por ejemplo, el color medio o la posición del centroide. Estos *trackers* dependen de una etapa anterior que detecta candidatos como *blobs* (por ejemplo, sustracción del fondo). Su principal ventaja es su alta eficiencia cuando sigues un objetivo con una cámara estacionaria u objetivos ampliamente separados y su escasa complejidad que permite el desarrollo de sistemas en tiempo real [59]. Sin embargo, algunas complicaciones pueden surgir cuando los objetivos se mueven cerca de la cámara.

***Template:*** Los *trackers* de tipo *template* describen una característica (por ejemplo, el valor de intensidad) del objetivo. Hay dos enfoques dependiendo si el *template* es rígido o deformable. El seguimiento por medio de un *template* rígido [60] supone que la información del objetivo no va a cambiar en *frames* consecutivos. Por lo tanto, métricas simples se pueden utilizar para localizar el objetivo (por ejemplo, la correlación de intensidad). Para el seguimiento por *template* deformable [61], se emplea un modelo de movimiento paramétrico para definir el movimiento del objetivo y, por lo tanto, aplican transformaciones al *template* del objetivo. El principal problema de este enfoque es el requisito de una etapa de aprendizaje que podría no ser posible debido a la disponibilidad de los datos.

***Kernel:*** Los métodos de histograma de núcleo básicamente utilizan un núcleo ponderado, así como un histograma para representar el objetivo. Estos *trackers* se pueden considerar como un punto medio entre los *trackers* de *blob* y los *trackers* de *template*. Los algoritmos disponibles difieren en varios aspectos, como el objetivo seguido, las características seleccionadas y el método para modelar el movimiento del objetivo. MeanShift [51] es uno de los *trackers* más populares de tipo *kernel*.

## 2.4 Seguimiento de múltiples personas basado en detección

Nuevas cámaras de video se instalan diariamente en todo el mundo para vigilancia o para una multitud de otros fines. Mientras esto sucede, se vuelve cada vez más importante desarrollar métodos que procesan dichos flujos de datos de forma automática y en tiempo real, reduciendo el esfuerzo manual que aún se requiere para el análisis de video. De particular interés para muchas aplicaciones es el comportamiento de las personas, por ejemplo, para la seguridad del tráfico, vigilancia, o el análisis deportivo. Como la mayoría de las tareas en los niveles semánticamente más altos se basan en la información de trayectoria, es crucial detectar robustamente y seguir a las personas en escenas dinámicas y complejas del mundo real. Sin embargo, la mayoría de métodos existentes de seguimiento todavía se limitan a escenarios de aplicaciones especiales. Requieren ya sea de entrada multi-cámara, conocimiento específico de la escena, un fondo estático, o información de profundidad, o no son apropiados para el procesamiento en línea.

El problema de la detección y el seguimiento de forma automática de un número variable de objetivos en escenas complejas son muy desafiantes debido a que existen muchas fuentes de incertidumbre para localizar las ubicaciones de los objetivos, como la medición de ruido, *clutter*, cambio de fondo y oclusiones significativas.

Con el fin de hacer frente a las dificultades descritas anteriormente, los enfoques de seguimiento por detección (*tracking-by-detection*) se han convertido cada vez más populares, impulsados por los últimos avances en los algoritmos de detección de objetos. Tales métodos implican la aplicación continua de un algoritmo de detección en *frames* individuales y la asociación de detecciones a través de los *frames*. En contraste con *trackers* basados en el modelado de fondo, son generalmente robustos en cambios de fondo y movimientos de la cámara.

El principal reto al utilizar un detector de objetos para realizar el seguimiento es que la salida del detector no es fiable y es escasa, es decir, los detectores sólo entregan un conjunto discreto de respuestas, y por lo general, producen falsos positivos y falsos negativos (*missing detections*). Por lo tanto, el problema resultante de asociación entre detecciones y objetivos es complicado. Varios algoritmos recientes se enfrentan a este problema mediante la optimización de tareas de detección en una gran ventana temporal en un paso *offline* [3][62][63][64]. Utilizan la información de *frames* futuros y localizan los objetivos en el *frame* actual con un retardo temporal o después de que la secuencia completa haya sido observada. Por el contrario, los métodos secuenciales de Monte Carlo ofrecen un marco para la representación de la incertidumbre de seguimiento de una forma causal. Sólo teniendo en cuenta la información de *frames* pasados, tales enfoques son más adecuados para aplicaciones de tiempo crítico, en línea.

Muchos métodos dependen de la sustracción de fondo de una o varias cámaras estáticas para llevar a cabo el seguimiento [62][65][66][67][68]. Sin embargo, como se ha comentado anteriormente, los recientes avances en la detección de objetos ha estimulado el interés por combinar el seguimiento y la detección. A diferencia de los enfoques de seguimiento basado en la asociación de datos, que vinculan las respuestas de detección a trayectorias mediante la optimización global basada en la posición, el tamaño y la similitud de apariencia [3][62][63][64][69][70][71] la combinación de los detectores de objeto y los resultados de filtrado de partículas resulta en algoritmos que son más adecuados para aplicaciones de tiempo crítico, en línea. Para este fin, [30] combina el algoritmo de [72]

con un detector de objetos *boosted* (mejorado). [73] amplia este Filtro de Partículas mejorado utilizando conjuntos de partículas independientes para cada objetivo de tal forma que se aumenta la robustez de múltiples objetivos. Además, para manejar oclusiones con mayor robustez, otros investigadores utilizan la información 3D [24][74], detectores entrenados por partes individuales del cuerpo [11], aplican modelos de movimiento de aplicación específica [75], o se basan en la salida intermedia proporcionada por el detector [4].

Por otro lado, hay investigadores que realizan el seguimiento mediante la explotación de un clasificador entrenado para distinguir entre el objeto y el fondo [76][77]. Del mismo modo, existen enfoques similares que aplican clasificadores con diferentes umbrales de confianza [78][79] o acumulan probabilidades de detección temporalmente [80][81].

Utilizar *trackers* independientes requiere la solución de un problema de asociación de datos para asignar detecciones a objetivos. Enfoques clásicos incluyen el *Joint Probabilistic Data Association Filter* (JPDAF) [8] y el *Multi Hypotheses Tracking* (MHT) [9]. MHT considera múltiples asociaciones posibles a lo largo de varios instantes de tiempo, pero su complejidad generalmente limita el análisis a sólo unos pocos pasos. JPDAFs, por otro lado, intenta hacer la mejor asignación posible en cada instante de tiempo considerando conjuntamente todas las posibles asociaciones entre los objetivos y detecciones, a costa de un aumento exponencial de la complejidad. Alternativamente, el algoritmo Húngaro [10] se puede utilizar para encontrar la mejor asignación posible de pares *detection-tracker* en un tiempo de ejecución que es cúbico con el número de objetivos. En la práctica, un algoritmo de tipo *greedy* es, sin embargo, a menudo suficiente, como el señalado por [4] y [11].

## 2.5 Evaluación

En esta sección se describen los diferentes *datasets* públicos y métricas existentes en el estado del arte que se utilizan para llevar a cabo una evaluación de los algoritmos de seguimiento múltiples personas.

### 2.5.1 Datasets

En este apartado se describen los *datasets* públicos existentes para llevar a cabo el proceso de evaluación de los resultados del sistema de seguimiento de múltiples personas.

#### 2.5.1.1 PETS

PETS (<http://www.cvg.rdg.ac.uk/slides/pets.html>) es el *dataset* más extendido en la actualidad. Desde el año 2000 establece una nueva base de datos cada año proponiendo retos diferentes a la comunidad investigadora. Con los algoritmos proporcionados los investigadores pueden probar o desarrollar nuevos algoritmos. Los mejores se presentan en la conferencia que se celebra cada año.

Dado que la cantidad de datos es amplia y cubre situaciones reales, estas bases de datos son las más utilizadas y son consideradas casi un estándar de facto. A pesar de esto, es

importante decir que las bases de datos de PETS no son ideales. Una de sus desventajas es el hecho de que desde PETS se convirtió en un proyecto de vigilancia, los retos propuestos se centran en aplicaciones de alto nivel de este campo, dejando a un lado el enfoque de seguimiento. Por lo tanto, algunas cuestiones importantes (como los cambios de escala del objetivo) no se consideran.

Para la realización de este proyecto, se ha utilizado la base de datos PETS2009 que se describe a continuación.

### 2.5.1.1.1 PETS2009

El *dataset* PETS2009 contiene diferentes escenarios que fueron grabados desde múltiples cámaras. El principal objetivo de este *dataset* es utilizar sistemas nuevos o existentes para:

- Estimación de la densidad en multitudes.
- Seguimiento de personas.
- Detección de eventos en multitudes.

Para ello, contiene un conjunto de secuencias de entrenamiento para cada cámara y tres conjuntos de secuencias de test, también para cada cámara. Los formatos para cada conjunto son imágenes en JPEG y no se dispone del *ground-truth* para el seguimiento. En la Figura 2-7 se muestra un ejemplo de un *frame* de cada cámara.



**Figura 2-7: Ejemplo de las 8 perspectivas disponibles en PETS2009.**

### 2.5.1.2 CAVIAR

El objetivo principal del *dataset* CAVIAR (<http://homepages.inf.ed.ac.uk/rbf/CAVIAR/>) es abordar la siguiente pregunta científica: ¿pueden las descripciones de imágenes locales de sensores de fovea u otro tipo, seleccionados mediante un proceso de atención visual jerárquico y guiados y procesados utilizando el conocimiento contextual del objeto, función y escena, mejorar los procesos de reconocimiento basados en imágenes?. Por tanto, uno de sus objetivos principales radica en la necesidad de representar y reconocer

objetos, contextos y situaciones. Varios métodos han sido investigados con el fin de abordar estas cuestiones en diferentes áreas.

Este *dataset* incluye secuencias de personas caminando solas, encuentro con otras, personas entrando y saliendo de tiendas, luchando, desmayándose y dejando algún paquete en algún lugar público. Todas las secuencias fueron grabadas con un amplio ángulo de la lente de la cámara y algunos escenarios fueron filmados con dos puntos de vista deferentes (sincronizados *frame a frame*). Está formado por dos conjuntos de secuencias:

- INRIA: Contiene 6 conjuntos de secuencias diferentes.
- Centro comercial en Portugal: Contiene 11 conjuntos de secuencias y 6 escenarios diferentes.

Para la realización de este proyecto se han utilizado únicamente las secuencias incluidas en el segundo conjunto de este dataset (Centro Comercial en Portugal). Los formatos para cada conjunto son imágenes en JPEG y vídeos en MJPEG y se dispone del *ground-truth* para el seguimiento en formato XML. En la Figura 2-8 se muestra un ejemplo de algunos *frames* de este escenario.



Figura 2-8: Ejemplo de algunos *frames* de CAVIAR.

### 2.5.1.3 VISOR

El *Video Surveillance Online Repository* (<http://www.openvisor.org/>) es una extensa base de datos que contiene un amplio conjunto de datos multimedia y las anotaciones correspondientes. El repositorio se ha concebido como una herramienta de apoyo a diferentes proyectos de investigación. Algunos vídeos están disponibles al público, sin embargo, la mayoría de ellos están restringidos y sólo se pueden ver después de realizar un registro. Los vídeos de la base de datos cubren una amplia gama de escenarios y situaciones, incluyendo vídeos para el reconocimiento de acciones humana, vídeos al aire libre para la detección de rostros, vídeos de interior para el seguimiento de personas con oclusiones en interior, videos para el reconocimiento humano, vídeos para la detección de vehículos y vigilancia del tráfico, etc.

Para la realización de este proyecto, se ha utilizado la sección del *dataset* que contiene los vídeos para el seguimiento de personas con oclusiones en interior la cual se explica a continuación.

#### 2.5.1.3.1 Vídeos para el seguimiento de personas con oclusiones en interior

Este conjunto de datos incluye varios vídeos con una amplia gama de oclusiones causadas por objetos o personas en la escena. Todos ellos incluyen anotaciones base y algunos también incluyen anotaciones automáticas. Está formado por seis secuencias en formato MJPEG y no se dispone de *ground-truth* para el seguimiento. En la Figura 2-9 se muestra un ejemplo de algunos *frames* del escenario incluido.



Figura 2-9: Ejemplo de algunos *frames* de VISOR.

#### 2.5.1.4 iLIDS

El *Imagery Library for Intelligent Detection Systems* (<http://www.eecs.qmul.ac.uk/~andrea/avss2007.html>) se incluyó en la Conferencia AVSS del 2007. Este *dataset* se utiliza para detectar el abandono de objetos en estaciones de metro y para detectar vehículos estacionados ilegalmente.

La base de datos está formada por siete secuencias, 3 de ellas son utilizadas para la detección de objetos abandonados y 4 de ellas para la detección de vehículos estacionados ilegalmente. Las secuencias poseen complejidad creciente para cada escenario. Adicionalmente, existen dos vídeos adicionales de larga duración y elevada complejidad para cada escenario. El formato de las secuencias se trata de imágenes JPEG y no dispone de *ground-truth* para el seguimiento.

Para la realización de este proyecto se han utilizado únicamente las secuencias incluidas para la detección de objetos abandonados en estaciones de metro. En la Figura 2-10 se muestra un ejemplo de algunos *frames* de este escenario.



Figura 2-10: Ejemplo de algunos *frames* de i-LIDS.

#### 2.5.2 Métricas

En este apartado se describen las métricas existentes para llevar a cabo una evaluación de los algoritmos de seguimiento de múltiples objetivos. Para medir el rendimiento se puede operar a nivel de *frame* o de secuencia ya sea considerándose trayectorias individuales o todas las trayectorias. Se pueden identificar tres categorías [82] de medidas de evaluación del seguimiento de múltiples objetivos:

- Evaluación PAP (*Point-based Assignment and Position-based evaluation*).
- Evaluación RAP (*Region-based Assignment and Position-based evaluation*).
- Evaluación RAS (*Region-based Assignment and Size-based evaluation*).

Estas tres categorías se explican después de definir la notación que se ha utilizado para definir cada una de las métricas que se incluyen dentro de dichas categorías.

### 2.5.2.1 Notación

En este apartado se describe la notación que se ha utilizado para definir cada una de las métricas que se incluyen dentro de las tres categorías de medidas de evaluación existentes para el seguimiento de múltiples objetivos.

Se define  $X_{k,j}$  como el estado estimado del objetivo  $j$  mediante un *tracker* en el *frame*  $k$  de la siguiente forma:

$$X_{k,j} = (x_{k,j}, y_{k,j}, A_{k,j}, l_j) \quad \text{Ecuación 1}$$

donde  $(x_{k,j}, y_{k,j})$  define la posición del objetivo,  $A_{k,j}$  indica su información de región (por ejemplo, *bounding box*, elipse, etc.),  $l_j$  es el ID del objetivo y  $k = 1, \dots, K$  donde  $K$  es el número de *frames* en la secuencia de vídeo. En el caso de objetivos basados en puntos, el estado estimado del objetivo  $j$  en el *frame*  $k$  no contiene  $A_{k,j}$  y se denota como  $X'_{k,j}$ .

Por tanto, se denomina a  $\mathbf{X}_k$  como el conjunto de los estados estimados de múltiples objetivos, es decir:

$$\mathbf{X}_k = \{X_{k,1}, \dots, X_{k,j}, \dots, X_{k,u_k}\} \quad \text{Ecuación 2}$$

donde  $u_k = |\mathbf{X}_k|$  es el número de objetivos estimados en el *frame*  $k$ .

La trayectoria  $\mathbf{x}_j$  del objetivo  $j$  se define como una secuencia de estados en el tiempo:

$$\mathbf{x}_j = \{X_{k,j}\}_{k=k_{ini}^j}^{k_{end}^j} \quad \text{Ecuación 3}$$

donde  $k_{ini}^j$  y  $k_{end}^j$  denotan los números de *frame* inicial y final de  $\mathbf{x}_j$ , respectivamente, y adicionalmente  $K_j$  es el número de *frames* totales de  $\mathbf{x}_j$ .

Por tanto, se denomina  $\chi$  como el conjunto que contiene todas las trayectorias estimadas en la secuencia, es decir:

$$\chi = \{\mathbf{x}_j\}_{j=1}^U \quad \text{Ecuación 4}$$

donde  $U$  define el número de trayectorias estimadas.

De forma similar,  $\bar{X}_{k,i}$ ,  $(\bar{x}_{k,i}, \bar{y}_{k,i}, \bar{A}_{k,i}, \bar{l}_i)$ ,  $\bar{\mathbf{X}}_k$ ,  $v_k$ ,  $\bar{X}'_{k,i}$ ,  $\bar{x}_i$ ,  $\bar{k}_{ini}^i$ ,  $\bar{k}_{end}^i$ ,  $\bar{K}_i$ ,  $\bar{\chi}$  y  $V$  son las correspondientes anotaciones del *ground-truth* para  $X_{k,j}$ ,  $(x_{k,j}, y_{k,j}, A_{k,j}, l_j)$ ,  $\mathbf{X}_k$ ,  $u_k$ ,  $X'_{k,j}$ ,  $\mathbf{x}_j$ ,  $k_{ini}^j$ ,  $k_{end}^j$ ,  $K_j$ ,  $\chi$  y  $U$ , respectivamente.

### 2.5.2.2 Evaluación PAP

Las medidas PAP utilizan una asignación basada en puntos y evalúan sólo la posición del objetivo, sin tener en cuenta cambios temporales de tamaño. Ejemplos de medidas PAP incluyen *Object Tracking Error* (OTE), la métrica basada en la distancia *Wasserstein's*, *Tracker Detection Rate* (TRDR), *False Alarm Rate* (FAR), *Track Detection Rate* (TDR) y *Track Fragmentation* (TF).

#### 2.5.2.2.1 Object Tracking Error (OTE)

OTE [83] calcula la distancia de posición promedio entre los pares del *ground-truth* y la trayectoria estimada. La asignación asocia una trayectoria estimada con la trayectoria del *ground-truth* que minimiza la distancia media Euclídea a través de sus *frames* comunes. Para cada par asociado  $t$ , su  $OTE_t$  se calcula como:

$$OTE_t = \frac{1}{\hat{K}_t} \sum_{k=\hat{k}_{ini}^t}^{\hat{k}_{end}^t} \sqrt{(\hat{x}_{k,t} - x_{k,t})^2 + (\hat{y}_{k,t} - y_{k,t})^2} \quad \text{Ecuación 5}$$

donde  $\hat{K}_t = \hat{k}_{end}^t - \hat{k}_{ini}^t$  es el número de *frames* que son comunes tanto en el *ground-truth* como en las trayectorias estimadas y  $\hat{k}_{ini}^t$  y  $\hat{k}_{end}^t$  denotan los números de *frame* inicial y final, respectivamente, del par  $t$ .

#### 2.5.2.2.2 Métrica basada en la distancia Wasserstein's

La métrica basada en la distancia *Wasserstein's* [84],  $W_p(\bar{\mathbf{X}}_k, \mathbf{X}_k)$  calcula la precisión del seguimiento de múltiples objetivos como:

$$W_p(\bar{\mathbf{X}}_k, \mathbf{X}_k) = \min_{\mathcal{C}} (\sum_{j=1}^{u_k} \sum_{i=1}^{v_k} C_{j,i}^k d(X'_{k,j}, \bar{X}'_{k,i})^p)^{1/p} \quad \text{Ecuación 6}$$

donde  $d(\cdot)^p$  denota la norma  $p$  con  $p \in [1, \infty)$ ,  $u_k$  es el número de objetivos estimados,  $v_k$  es el número de objetivos del *ground-truth*, y  $\mathcal{C}$  es la matriz traspuesta que define el coste de asociación entre todos los posibles pares de trayectorias estimadas y del *ground-truth*. Las asociaciones que minimizan el coste total se determinan utilizando los algoritmos Húngaro [10] y Munkres [85].

#### 2.5.2.2.3 Tracker Detection Rate (TRDR), False Alarm Rate (FAR), Track Detection Rate (TDR) y Track Fragmentation (TF)

TRDR, FAR y TDR [83] evalúan la exactitud utilizando los verdaderos positivos (*true positives*) y falsos positivos (*false positives*) determinados con el criterio de coincidencia. Aunque estas medidas utilizan la información del tamaño del objetivo en la evaluación, se trata de medidas PAP porque no evalúan los cambios de tamaño del objetivo con el tiempo. Para TRDR, FAR y TDR, la asignación entre las trayectorias del *ground-truth* y las estimadas se resuelven igual que para OTE.

TRDR cuantifica el rendimiento general en el *frame*  $k$  como el ratio entre el número de objetivos seguidos correctamente (*true positives*),  $|\widehat{TP}_k|$ , y el número de objetivos del *ground-truth*  $v_k$ , es decir:

$$TRDR_k = \frac{|\widehat{TP}_k|}{v_k} \quad \text{Ecuación 7}$$

Una estimación se considera verdadero positivo si el centroide del *bounding box* del *ground-truth* se encuentra dentro del *bounding box* estimado. Si ningún centroide del *ground-truth* coincide con el *bounding box* estimado, la estimación se considera como un falso positivo.

FAR cuantifica el rendimiento del seguimiento en el *frame*  $k$  como la relación entre el número de objetivos incorrectamente seguidos (falsos positivos),  $|\widehat{FP}_k|$ , y la suma de objetivos seguidos correctamente e incorrectamente,  $|\widehat{TP}_k| + |\widehat{FP}_k|$ :

$$FAR_k = \frac{|\widehat{FP}_k|}{|\widehat{TP}_k| + |\widehat{FP}_k|} \quad \text{Ecuación 8}$$

TDR cuantifica el rendimiento de seguimiento a nivel de trayectoria como la relación entre el número de verdaderos positivos en la trayectoria estimada  $x_j$ ,  $TP_j$ , y el número de *frames* donde existe la correspondiente trayectoria del *ground-truth*  $x_i$ ,  $\bar{K}_i$ :

$$TDR_i = \frac{TP_j}{\bar{K}_i} \quad \text{Ecuación 9}$$

La evaluación de la consistencia de los IDs de los objetivos se proporciona en la forma de TF:

$$TF_i = |IDC_i| \quad \text{Ecuación 10}$$

donde  $|IDC_i|$  es el número de cambios de ID con respecto a la trayectoria  $i$  del *ground-truth*, medido como el número de veces que la trayectoria  $i$  del *ground-truth* se asocia con diferentes trayectorias estimadas.

### 2.5.2.3 Evaluación RAP

Las medidas RAP utilizan una asignación basada en región y proporcionan una evaluación basada en posición. Ejemplos de medidas RAP incluyen las trayectorias emparejadas  $\widehat{TP}$ ,  $\widehat{FP}$  y  $\widehat{FN}$ .

El cálculo de las trayectorias emparejadas  $\widehat{TP}$ ,  $\widehat{FP}$  y  $\widehat{FN}$  [86] se basa en el solapamiento espacial y temporal entre las trayectorias estimada y del *ground-truth* y implica realizar la asignación de forma implícita. Si la trayectoria estimada  $j$  se solapa con cualquier trayectoria  $i$  del *ground-truth* tanto temporalmente como espacialment, entonces, la estimación se considera un emparejamiento  $\widehat{TP}$ . Una superposición espacial se logra en un *frame* cuando el centroide de la trayectoria estimada  $j$  coincide con el correspondiente *bounding box* de la trayectoria  $i$  del *ground-truth*. A nivel de trayectoria, se mide para cada trayectoria del *ground-truth* el porcentaje de *frames* que tienen coincidencia entre los

*bounding boxes* estimados y del *ground-truth*. Para un emparejamiento  $\widehat{TP}$ , el solape temporal,  $\bar{O}_{fp}$ , entre la trayectoria estimada  $j$  y la correspondiente trayectoria  $i$  del *ground-truth* se define como:

$$\bar{O}_{tp} = \frac{N_{i,j}^{ov}}{K_i} \quad \text{Ecuación 11}$$

donde  $N_{i,j}^{ov}$  es el número de *frames* concurrentes entre la trayectoria  $i$  del *ground-truth* y la trayectoria  $j$  estimada, es decir, los *frames* donde  $i$  y  $j$  existen.

Si la superposición espacial o temporal de la trayectoria  $j$  estimada con cualquier trayectoria  $i$  del *ground-truth* es menor que un umbral  $\bar{\tau}_2$ , la estimación se considera un  $\widehat{FP}$ . Para un  $\widehat{FP}$ , el solapamiento temporal  $\bar{O}_{fp}$ , entre la trayectoria estimada  $j$  y la correspondiente trayectoria  $i$  del *ground-truth* se define como:

$$\bar{O}_{fp} = \frac{N_{i,j}^{ov}}{K_j} \quad \text{Ecuación 12}$$

Teniendo en cuenta todas las trayectorias estimadas, si la superposición espacial o temporal de la trayectoria  $i$  del *ground-truth* con cualquier trayectoria estimada  $j$  es menor que un umbral  $\bar{\tau}_2$ , la estimación se considera un  $\widehat{FN}$ . Para un  $\widehat{FN}$ , el solapamiento temporal  $\bar{O}_{fn}$  se calcula como para  $\bar{O}_{tp}$ .

#### 2.5.2.4 Evaluación RAS

Las medidas RAS utilizan una asignación basada en región y proporcionan una evaluación del seguimiento que también tiene en cuenta los cambios del tamaño del objetivo en el tiempo. Ejemplos de medidas RAS incluyen *Correct Detected Track* (CDT), *False Alarm Track* (FAT), *Track Detection Failure* (TDF), *Multiple Object Tracking Precision* (MOTP), *Multiple Object Detection Accuracy* (MODA), *Normalized MODA* y *Multiple Object Tracking Accuracy* (MOTA).

##### 2.5.2.4.1 Correct Detected Track (CDT), False Alarm Track (FAT) y Track Detection Failure (TDF)

CDT, FAT y TDF [87] son conceptualmente similares a las trayectorias  $\widehat{TP}$ ,  $\widehat{FP}$  y  $\widehat{FN}$ , respectivamente. Sin embargo, a diferencia de ellos, el solapamiento espacial se define utilizando el número de píxeles comunes entre los *boundig boxes* del *ground-truth* y el estimado. Esto implica que CDT, FAT y TDF incluyen también las variaciones del tamaño del objetivo en la evaluación.

##### 2.5.2.4.2 Multiple Object Tracking Precision (MOTP), Multiple Object Detection Accuracy (MODA), Normalized MODA y Multiple Object Tracking Accuracy (MOTA).

Para MOTP, MODA, *Normalized MODA* y MOTA [88] se realiza una asignación uno a uno a nivel de *frame* entre las trayectorias del *ground-truth* y las estimadas basada en la

maximización de los valores de solapamiento espacial (calculado como para CDT, FAT y TDF) entre pares utilizando el algoritmo Húngaro [10].

MOTP es una medida espacio-temporal que calcula la cantidad de solapamiento entre las trayectorias estimada y la del *ground-truth* mediante:

$$MOTP = \frac{\sum_{t=1}^{n_m} \sum_{k=\widehat{k}_{ini}^t}^{\widehat{k}_{end}^t} \frac{|\bar{A}_k^t \cap A_k^t|}{|\bar{A}_k^t \cup A_k^t|}}{\sum_{k=1}^K n_m^k} \quad \text{Ecuación 13}$$

donde  $n_m$  es el número de pares asociadas de trayectorias estimadas y del *ground-truth*,  $|\bar{A}_k^t \cap A_k^t|$  es el número de píxeles comunes en  $\bar{A}_k^t$  y  $A_k^t$ ,  $|\bar{A}_k^t \cup A_k^t|$  es el número de píxeles en  $\bar{A}_k^t \cup A_k^t$ , y  $n_m^k$  es el número de pares asociados de objetivos del *ground-truth* y estimados en el *frame*  $k$ . Los pares con un solapamiento mayor que un valor de umbral fijado  $\tau_0$  se consideran en el procedimiento de evaluación.

MODA $_k$  calcula el rendimiento del seguimiento en el *frame*  $k$  mediante la combinación de la información del número de estimaciones de falsos positivos  $\widehat{FN}_k$  y el número de estimaciones de falsos positivos  $\widehat{FP}_k$ :

$$MODA_k = 1 - \frac{|\widehat{FN}_k| + |\widehat{FP}_k|}{v_k} \quad \text{Ecuación 14}$$

donde  $\widehat{FN}_k$  y  $\widehat{FP}_k$  se determinan mediante la comparación de la cantidad de solapamiento entre los objetivos correspondientes del *ground-truth* y los estimados con el umbral  $\tau_0$ .

Una formulación de MODA a nivel de secuencia, la *Normalized MODA* (N-MODA), se define como:

$$N - MODA_k = 1 - \frac{\sum_{k=1}^K (|\widehat{FN}_k| + |\widehat{FP}_k|)}{\sum_{k=1}^K v_k} \quad \text{Ecuación 15}$$

A diferencia de MODA, MOTA es una medida a nivel de secuencia que evalúa el rendimiento del seguimiento añadiendo también la información del número de cambios de ID ( $|IDS_k|$ ) en cada *frame*, además de  $|\widehat{FN}_k|$  y  $|\widehat{FP}_k|$ . Las contribuciones se acumulan a través de la secuencia y se normalizan de la siguiente forma:

$$MOTA = 1 - \frac{\sum_{k=1}^K (|\widehat{FN}_k| + |\widehat{FP}_k| + |IDS_k|)}{\sum_{k=1}^K v_k} \quad \text{Ecuación 16}$$

donde  $|\widehat{FN}_k|$  y  $|\widehat{FP}_k|$  se determinan como en MODA y  $|IDS_k|$  cuenta el número de cambios de ID correspondientes a las trayectorias del *ground-truth* en cada *frame*.

## 2.6 Conclusiones

Durante este capítulo se ha realizado un estudio detallado del estado del arte del seguimiento de múltiples personas en vídeo basado en detección. Para ello, se han estudiado los algoritmos existentes en la literatura actual de detección de personas (sección

2.2), de seguimiento de objetos (sección 2.3) y aquellos que combinan el seguimiento con la detección para múltiples personas (sección 2.4). Además, se han analizado los distintos *datasets* públicos y métricas existentes actualmente en el estado del arte que se utilizan para llevar a cabo una evaluación del algoritmo de seguimiento múltiple (sección 2.5).

En la sección 2.2, se han descrito las diferentes tareas de procesamiento que implican la detección automática de personas realizándose una clasificación completa de la detección de personas independientemente de su aplicación de vídeo-vigilancia posterior. Las tareas críticas en cualquier algoritmo de detección de personas son la generación o extracción de las hipótesis iniciales de objetos a ser considerados personas en la escena y el modelo de persona utilizado para clasificar dichas hipótesis (sección 2.2.2). El enfoque de la detección de objetos tiene una gran influencia sobre los resultados. La segmentación es una técnica simple y poderosa de extracción de objetos, pero que posee muchas dificultades y limitaciones en entornos complejos. En contraste, la búsqueda exhaustiva es más robusta a la rotación y a cambios de escala y de pose incluso en entornos complejos, pero tiene la complejidad de la adición de muchos falsos positivos, además de un coste computacional superior. Por otro lado, el modelo de persona elegido para clasificar los objetos candidatos iniciales a ser persona determina la robustez del algoritmo a las variaciones de la persona y oclusiones. Los modelos simples (holísticos) basados sólo en modelos de movimiento o de apariencia son menos robustos a variaciones de personas y oclusiones, mientras que los modelos basados en parte añaden complejidad al algoritmo, pero son mucho más robustos a las variaciones de la gente y oclusiones.

En la sección 2.3 se ha sintetizado la información más relevante sobre el seguimiento de objetos en vídeo. Es una tarea complicada debido a la alta variabilidad y complejidad de los datos a analizar y, por tanto, se requiere seguir una serie de pasos para el diseño de un algoritmo de seguimiento, tales como la extracción de características, la representación del objetivo y la propagación del modelo del objetivo en el tiempo. Las características elegidas (ya sean de bajo nivel, nivel medio o alto nivel) para representar el objetivo tienen que identificar singularmente el objetivo del fondo de la escena u otros objetos. Su elección dependerá de la situación en la que se utilice el algoritmo. Además, las características del objetivo tienen que estar representadas adecuadamente por medio de un modelo (por ejemplo, básico, aproximaciones de área o volumen, etc.), de tal forma que se pueda proporcionar un seguimiento preciso en el vídeo. Por otro lado, para determinar la localización de un objetivo en una determinada posición de la imagen se dispone de los métodos de búsqueda exhaustiva y búsqueda selectiva. El primero de ellos, busca en todas las regiones del área de búsqueda con el objetivo de minimizar la función de coste que define la similitud entre el objetivo a ser seguido y las características observadas en el área de búsqueda, mientras que el segundo, utiliza un modelo de movimiento del objetivo para llevar a cabo una etapa de predicción que proyecta el estado al instante actual y una etapa de actualización que comprueba las similitudes entre las predicciones y el modelo del objetivo.

En la sección 2.4 se han resumido los métodos existentes que combinan la detección y el seguimiento de múltiples personas. Este tipo de técnicas se han convertido en muy populares debido a los grandes avances en los algoritmos de detección de objetos. Tales métodos implican la aplicación continua de un algoritmo de detección en *frames* individuales y la asociación de detecciones a través de los *frames*. Muchos algoritmos realizan una optimización de las tareas de detección en un modo *offline* para hacer frente a los problemas de la salida no fiable del detector, mientras que los algoritmos basados en

métodos secuenciales de Monte Carlo sólo tienen en cuenta información de *frames* pasados de tal forma que son más adecuados para aplicaciones de tiempo crítico, en línea.

Por último, en la sección 2.5 se han analizado cada uno de los *datasets* públicos y métricas existentes en el estado del arte que permiten llevar a cabo una evaluación objetiva del rendimiento de los algoritmos de seguimiento que se desarrollan.

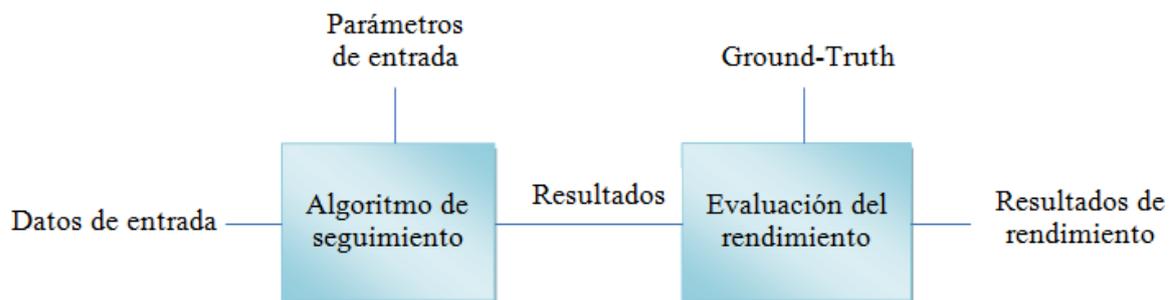
# 3 Protocolo de evaluación propuesto

## 3.1 Introducción

En este capítulo se presenta el protocolo de evaluación propuesto para evaluar el algoritmo de seguimiento de múltiples personas a desarrollar. En primer lugar, se indica la metodología de evaluación objetiva (sección 3.2) y subjetiva (sección 3.3) que se ha utilizado para medir el rendimiento del sistema. Seguidamente, se indica el *dataset* propuesto (sección 3.4) y las métricas utilizadas (sección 3.5) para llevar a cabo la evaluación del algoritmo de seguimiento. Por último, se han extraído una serie de conclusiones de cada una de las secciones que contiene el capítulo (sección 3.6).

## 3.2 Metodología de evaluación objetiva

La propuesta metodológica para evaluar objetivamente el sistema de seguimiento de múltiples personas a desarrollar se muestra en la Figura 3-1. Como se puede observar, se compone de dos etapas: algoritmo de seguimiento y evaluación del rendimiento.



**Figura 3-1: Metodología de evaluación objetiva propuesta para el sistema de seguimiento de múltiples personas**

### 3.2.1 Algoritmo de seguimiento

Este módulo lleva a cabo la ejecución del algoritmo de seguimiento que se quiere evaluar, con las siguientes entradas y salidas:

- Datos de entrada: Secuencia de vídeo que va a ser analizada.
- Configuración: Incluye todos los parámetros del sistema de seguimiento que el usuario puede ajustar manualmente (por ejemplo, umbrales del proceso de asociación).
- Resultados: Se genera un archivo con la información del seguimiento de cada uno de los objetivos.

Por tanto, la salida final del algoritmo de seguimiento genera un archivo con la información característica de cada uno de los objetivos en cada *frame* de la secuencia de

entrada. El formato de este archivo tiene que tener la siguiente estructura para que el módulo de evaluación del rendimiento funcione correctamente:

*frame ID<sub>objetivo</sub> numBlob x y w h*

donde *frame* es el número del fotograma anotado, *ID<sub>objetivo</sub>* es el identificador del objetivo, *numBlob* indica la posición de inclusión del objetivo en el archivo de anotación para el *frame* actual, *x* y *y* son las coordenadas de la esquina superior izquierda del *bounding box* del objetivo y *w* y *h* indican el ancho y la altura del mismo *bounding box*, respectivamente. Un ejemplo del formato del archivo de resultados se muestra en la Figura 3-2 en la que cada línea del archivo describe las coordenadas *x, y, w, h* de los objetivos 1 y 2 en los *frames* del 2 al 13:

2	1	1	119.000000	64.000000	36.000000	101.000000
3	1	1	119.300003	65.199997	35.400002	98.900002
4	1	1	120.699997	67.300003	34.000000	94.000000
5	1	1	120.699997	67.300003	34.000000	94.000000
6	1	1	120.699997	67.300003	34.000000	94.000000
7	1	1	120.000000	67.300003	34.000000	94.000000
8	1	1	123.099998	68.000000	34.000000	94.000000
8	2	2	174.000000	152.399994	47.099998	130.300003
9	1	1	124.500000	68.000000	34.000000	94.000000
9	2	2	176.100006	153.800003	45.000000	124.000000
10	1	1	125.900002	68.000000	34.000000	94.000000
10	2	2	176.800003	153.100006	45.000000	124.000000
11	1	1	126.599998	67.300003	34.000000	94.000000
11	2	2	177.500000	153.100006	45.000000	124.000000
12	1	1	127.300003	68.000000	34.000000	94.000000
12	2	2	178.199997	153.100006	45.000000	124.000000
13	1	1	128.000000	68.000000	34.000000	94.000000
13	2	2	178.899994	153.800003	45.000000	124.000000

**Figura 3-2: Ejemplo de algunas líneas del archivo generado tras la ejecución del algoritmo de seguimiento**

### 3.2.2 Evaluación del rendimiento

Este módulo lleva a cabo la evaluación del rendimiento del sistema de seguimiento a partir de los resultados obtenidos tras la ejecución del algoritmo de seguimiento y la métrica empleada. Sus entradas y salidas son:

- Resultados del seguimiento: Archivo de resultados como se ha descrito en el punto anterior.
- *Ground-Truth*: Este archivo contiene las anotaciones de la verdadera posición de cada uno de los objetivos de cada *frame*. El formato de línea de este archivo es el mismo que el del archivo de los resultados de seguimiento salvo que en este caso no se incluye el número de inclusión del *blob* (*numBlob*). Por tanto, está formado por seis columnas correspondientes al número del fotograma, el identificador del objetivo, las coordenadas de la esquina superior izquierda del *bounding box* del objetivo y su ancho y altura. Un ejemplo del formato del archivo de anotación del *ground-truth* se muestra en la Figura 3-3.

- Resultados de rendimiento: Resultados de rendimiento obtenidos para la métrica utilizada. Para ello, se compara el *ground-truth* con los resultados de seguimiento determinados.

5	1	323.000000	133.000000	70.000000	171.000000
5	2	375.000000	140.000000	91.000000	265.000000
6	1	322.000000	129.000000	68.000000	175.000000
6	2	375.000000	140.000000	91.000000	261.000000
7	1	320.000000	129.000000	68.000000	175.000000
7	2	379.000000	138.000000	91.000000	261.000000
8	1	320.000000	129.000000	68.000000	175.000000
8	2	379.000000	138.000000	91.000000	261.000000
9	1	318.000000	129.000000	68.000000	175.000000
9	2	379.000000	138.000000	91.000000	261.000000
10	1	316.000000	130.000000	68.000000	175.000000
10	2	382.000000	139.000000	91.000000	261.000000

Figura 3-3: Ejemplo de algunas líneas de un archivo de anotación del *ground-truth*.

### 3.3 Metodología de evaluación subjetiva

La propuesta metodológica para evaluar subjetivamente el sistema de seguimiento de múltiples personas a desarrollar se muestra en la Figura 3-4. Como se puede observar, a diferencia de la metodología de evaluación objetiva, se compone sólo de la etapa de ejecución del algoritmo de seguimiento.

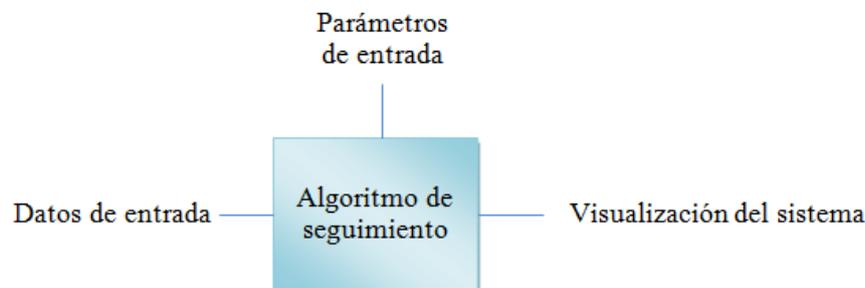


Figura 3-4: Metodología de evaluación subjetiva propuesta para el sistema de seguimiento de múltiples personas.

La salida visual del algoritmo de seguimiento consta de un *bounding box* para cada objetivo estimado, el cuál delimita el área de imagen que cubre de dicho objetivo, y de su identificador característico. Además, durante el desarrollo del seguimiento de un objetivo se muestra el recorrido que va siguiendo durante el transcurso de la escena. Para ello, se determina el centroide del *bounding box* estimado de cada uno de los objetivos y se almacena este valor en cada *frame* de la secuencia que se mantiene activo mediante la siguiente relación:

$$x_c = x_{bb} + \frac{w_{bb}}{2} \quad \text{Ecuación 17}$$

donde  $x_c$  es la coordenada  $x$  del centroide del *bounding box* del objetivo,  $x_{bb}$  es la coordenada  $x$  de la esquina superior izquierda del *bounding box* y  $w_{bb}$  es el ancho del *bounding box*.

$$y_c = y_{bb} + \frac{h_{bb}}{2}$$

**Ecuación 18**

donde  $y_c$  es la coordenada  $y$  del centroide del *bounding box* del objetivo,  $y_{bb}$  es la coordenada  $y$  de la esquina superior izquierda del *bounding box* y  $h_{bb}$  es la altura del *bounding box*.

En la Figura 3-5 se muestra un ejemplo de la visualización del *ground-truth* y del sistema final en el que se aprecia el seguimiento de un objetivo en una secuencia de vídeo.



**Figura 3-5:** Salida visual del *ground-truth* y del algoritmo de seguimiento de un objetivo.

### **3.4 Dataset**

La selección de los escenarios de prueba es uno de los pasos más importantes para el desarrollo de una evaluación del sistema. El principal problema en el seguimiento de múltiples personas radica en las oclusiones tanto parciales como totales que pueden aparecer a lo largo de la escena.

Para medir el rendimiento del algoritmo de seguimiento implementado se ha realizado un *dataset* a partir de los *datasets* públicos explicados en la sección 2.5 formado por seis niveles de complejidad. La ventaja principal de utilizar este *dataset* frente a otros del estado del arte es que los niveles se han diseñado de tal forma que la complejidad crece progresivamente desde niveles con secuencias simples en las que aparecen personas sin interactuar entre sí, hasta niveles con secuencias en las que aparecen múltiples personas interactuando entre ellas. En contraposición, alguno de los *datasets* públicos contienen un mayor número de secuencias que pueden cubrir más situaciones problemáticas.

A continuación, se describe cada uno de los niveles del *dataset* propuesto:

- Nivel 1: Este nivel está formado por tres secuencias de vídeo en las que aparecen dos personas sin interactuar entre sí, es decir, sin oclusiones.

	Secuencia 1	Secuencia 2	Secuencia 3
<b>Dataset público</b>	CAVIAR OneStopNoEnter1cor.mpg	CAVIAR OneStopNoEnter2cor.mpg	ETRI_od_A.mpg
<b>Frames</b>	141	150	750
<b>frames/s</b>	25	25	25

- Nivel 2: Este nivel está formado por 3 secuencias de vídeo en las que aparecen tres, cuatro o cinco personas sin interactuar entre sí, es decir, sin oclusiones.

	Secuencia 1	Secuencia 2	Secuencia 3
<b>Dataset público</b>	CAVIAR OneStopNoEnter2cor.mpg	PETS09 S2.L1	iLids AVSS_AB_Hard.avi
<b>Frames</b>	275	53	156
<b>frames/s</b>	25	10	25

- Nivel 3: Este nivel está formado por 5 secuencias de vídeo con dos, tres, cuatro o cinco personas, las cuales pueden presentar oclusiones parciales a lo largo del vídeo.

	Secuencia 1	Secuencia 2	Secuencia 3	Secuencia 4	Secuencia 5
<b>Dataset público</b>	iLids AVSS_AB_Easy.avi	iLids AVSS_AB_Easy.avi	CAVIAR ShopAssistant1cor.mpg	CAVIAR OneStopEnter2cor.mpg	PETS09 S2.L1
<b>Frames</b>	423	520	170	164	37
<b>frames/s</b>	25	25	25	25	10

- Nivel 4: Este nivel está formado por 3 secuencias de vídeo en las que aparecen dos o tres personas que se cruzan entre sí en algún momento de la secuencia, es decir, tiene lugar alguna oclusión total.

	Secuencia 1	Secuencia 2	Secuencia 3
<b>Dataset público</b>	CAVIAR OneLeaveShopReenter1cor.mpg	iLids AVSS_AB_Easy.avi	VISOR Indoor_tracking_video_3.avi
<b>Frames</b>	390	305	120
<b>frames/s</b>	25	25	25

- Nivel 5: Este nivel está formado por 2 secuencias de vídeo en las que aparecen siete personas como mínimo sin interactuar entre sí, es decir, sin oclusiones totales.

	Secuencia 1	Secuencia 2
<b>Dataset público</b>	PETS09 S2.L1	PETS09 S2.L1
<b>Frames</b>	73	85
<b>frames/s</b>	10	10

- Nivel 6: Este nivel está formado por 3 secuencias de vídeo en las que aparecen siete personas como mínimo interactuando entre sí, es decir, con oclusiones totales.

	Secuencia 1	Secuencia 2	Secuencia 3
<b>Dataset público</b>	PETS09 S2.L1	PETS09 S2.L1	PETS09 S2.L1
<b>Frames</b>	92	85	98
<b>frames/s</b>	10	10	10

Por otro lado, en la sección 2.5 se ha indicado los *datasets* públicos que contenían las anotaciones del *ground-truth* para cada una de sus secuencias. El *dataset* CAVIAR tenía dichas anotaciones en formato XML que no es compatible con el formato de archivo de entrada del módulo de evaluación del rendimiento descrito en la sección 3.2.2. Para aprovechar las anotaciones realizadas en formato XML de este *dataset* se ha realizado un *script* que adapta este formato al empleado en el módulo de evaluación del rendimiento. Por otro lado, el *dataset* PETS09 no poseía las anotaciones del *ground-truth* de sus secuencias. Sin embargo, [89] ha publicado dichas anotaciones también en formato XML como CAVIAR (<http://research.milanton.de/data.html>). Por tanto, para utilizar el *ground-truth* de [89] se ha realizado otro *script* que adapta también dicho formato al utilizado en el módulo de evaluación del rendimiento. Por último, para las secuencias de los demás *datasets* públicos utilizados (iLids, VISOR) que no contenían publicadas las anotaciones, se ha realizado el *ground-truth* utilizando la herramienta *Video Image Annotation Tool*.

A continuación, se incluye en la Figura 3-6 y en la Figura 3-7 algunos ejemplos que muestran las características de cada nivel.

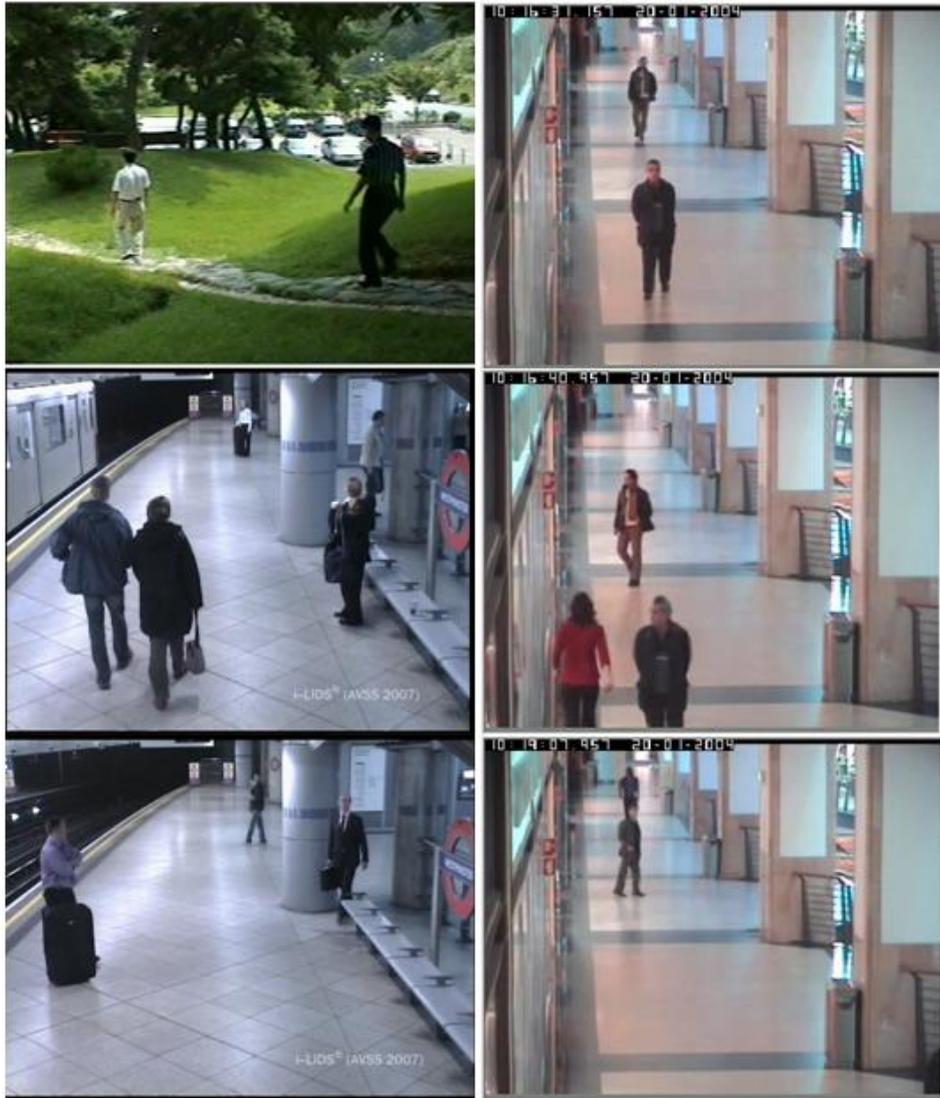
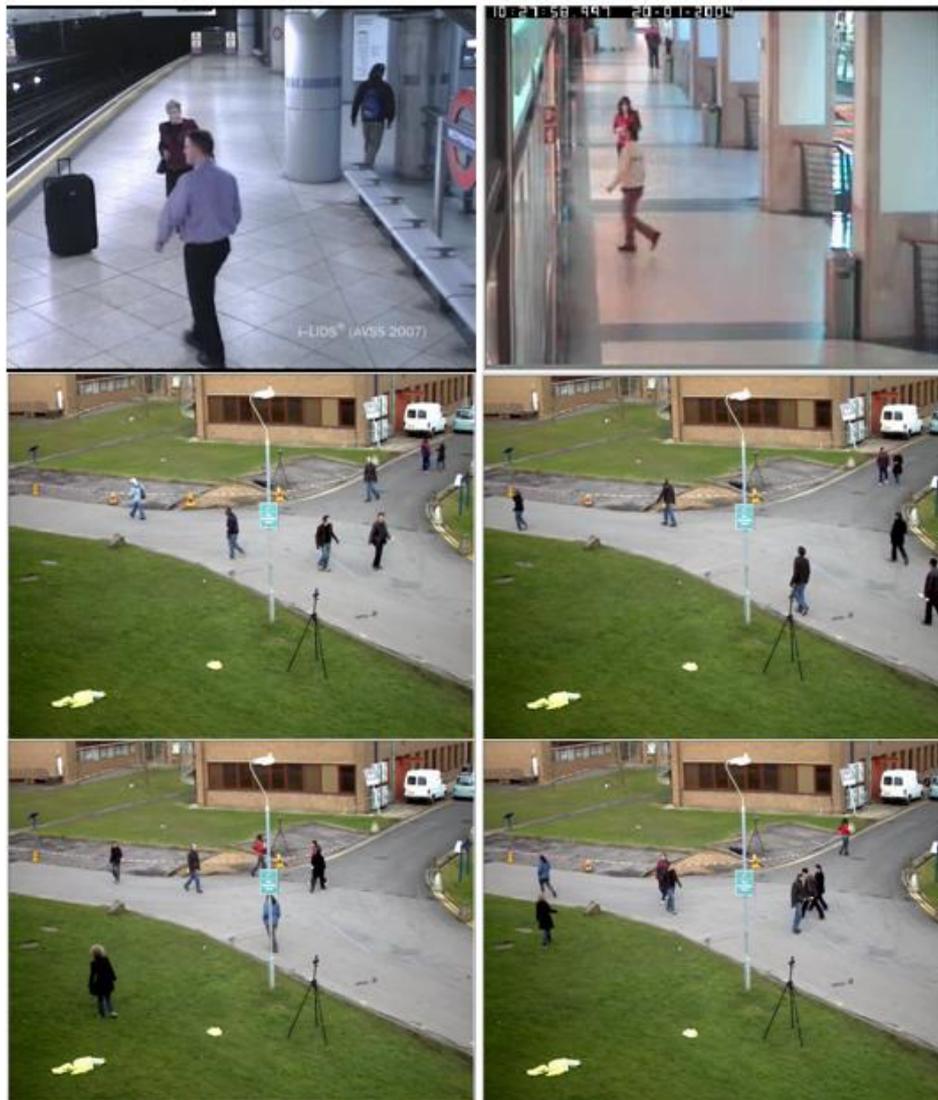


Figura 3-6: Ejemplo de algunos *frames* de alguna de las secuencias de los 3 primeros niveles del *dataset* propuesto (cada fila representa un nivel del *dataset* en orden creciente de dificultad).



**Figura 3-7:** Ejemplo de algunos *frames* de alguna de las secuencias de los 3 últimos niveles del *dataset* propuesto (cada fila representa un nivel del *dataset* en orden creciente de dificultad).

### 3.5 Métrica

El objetivo del seguimiento de múltiples personas es asignar un identificador único a cada persona que se mantiene constante a lo largo de la secuencia. Por lo tanto, teniendo en cuenta este objetivo, una métrica que mida el rendimiento del sistema debe ser capaz de determinar si el algoritmo de seguimiento es preciso en cuanto a la localización del objetivo y si el algoritmo es capaz de seguir el objetivo mientras se mueve a través de la escena incluso si es temporalmente ocluido (parcialmente o completamente).

Para evaluar el rendimiento del sistema de seguimiento de múltiples personas, se ha decidido emplear las métricas MOTP y MOTA [88] debido a que son muy utilizadas en la comunidad investigadora [4][62][65][90][91][92][93] y la mayoría de autores recomiendan su utilización a la hora de medir los resultados del algoritmo. A continuación, se describen las métricas que define [88]:

- **Multiple Object Tracking Precision (MOTP):**

El objetivo de esta métrica es calcular el solapamiento espacio-temporal entre los objetivos de salida del algoritmo de seguimiento y las anotaciones del *ground-truth*.

$$MOTP = \frac{\sum_{t=1}^{n_m} \sum_{k=\hat{k}_{ini}^t}^{\hat{k}_{end}^t} \frac{|\bar{A}_k^t \cap A_k^t|}{|\bar{A}_k^t \cup A_k^t|}}{\sum_{k=1}^K n_m^k} \quad \text{Ecuación 13}$$

donde  $n_m$  es el número de pares asociadas de trayectorias estimadas y del *ground-truth*,  $|\bar{A}_k^t \cap A_k^t|$  es el número de píxeles comunes en  $\bar{A}_k^t$  y  $A_k^t$ ,  $|\bar{A}_k^t \cup A_k^t|$  es el número de píxeles en  $\bar{A}_k^t \cup A_k^t$ , y  $n_m^k$  es el número de pares asociados de objetivos del *ground-truth* y estimados en el *frame*  $k$ .

- **Multiple Object Tracking Accuracy (MOTA):**

El objetivo de esta métrica es extraer la exactitud de la salida del sistema calculando el número de detecciones perdidas y falsos positivos, así como los cambios de ID que se producen a lo largo de la secuencia para cada una de las anotaciones del *ground-truth* disponibles.

$$MOTA = 1 - \frac{\sum_{k=1}^K (|\widehat{FN}_k| + |\widehat{FP}_k| + |IDS_k|)}{\sum_{k=1}^K v_k} \quad \text{Ecuación 16}$$

donde  $|\widehat{FN}_k|$ ,  $|\widehat{FP}_k|$ ,  $|IDS_k|$  indican el número de detecciones perdidas, falsos positivos y cambios de ID, respectivamente, y  $v_k$  es el número de objetivos del *ground-truth* en el *frame*  $k$ .

Por tanto, la métrica MOTA se puede interpretar como la combinación de tres ratios de error:

- Ratio de detecciones perdidas en la secuencia:  $\widehat{FN} = \frac{\sum_{k=1}^K |\widehat{FN}_k|}{\sum_{k=1}^K v_k}$
- Ratio de falsos positivos en la secuencia:  $\widehat{FP} = \frac{\sum_{k=1}^K |\widehat{FP}_k|}{\sum_{k=1}^K v_k}$
- Ratio de cambios de ID en la secuencia:  $IDS = \frac{\sum_{k=1}^K |IDS_k|}{\sum_{k=1}^K v_k}$

Para determinar si un objetivo de salida del sistema tiene una asociación con alguna anotación del *ground-truth* se determina el solapamiento entre dicho objetivo y la anotación del *ground-truth* más cercana. Si dicho solapamiento supera el 50% entonces se asocia el objetivo de salida a la anotación, mientras que si está por debajo de ese porcentaje entonces, se considerará como un falso positivo.

### **3.6 Conclusiones**

Durante este capítulo se ha presentado el protocolo de evaluación que se ha propuesto para evaluar el algoritmo de seguimiento de múltiples personas. Para ello, se han detallado las dos metodologías de evaluación objetiva (sección 3.2) y subjetiva (sección 3.3) utilizadas para medir el rendimiento del sistema, y después, se ha explicado el *dataset* que se ha propuesto y las métricas utilizadas para realizar la evaluación del algoritmo.

En la sección 3.2 se han explicado cada una de las etapas que componen la propuesta metodológica objetiva que se ha utilizado en este proyecto para la evaluación del algoritmo de seguimiento múltiple. Además, se ha detallado el formato de entrada y salida que tiene que tener cada uno de los módulos para que el sistema de evaluación funcione correctamente.

En la sección 3.3 se ha presentado la propuesta metodológica subjetiva que se ha utilizado para medir subjetivamente el rendimiento del algoritmo mediante la visualización del *bounding box* y de la trayectoria de cada uno de los objetivos presentes en el sistema.

Por último, en la sección 3.4 se ha explicado cada uno de los niveles de complejidad que contiene el *dataset* propuesto y en la sección 3.5 se ha indicado las métricas utilizadas para la evaluación objetiva del sistema.

## 4 Algoritmos seleccionados

---

### 4.1 Introducción

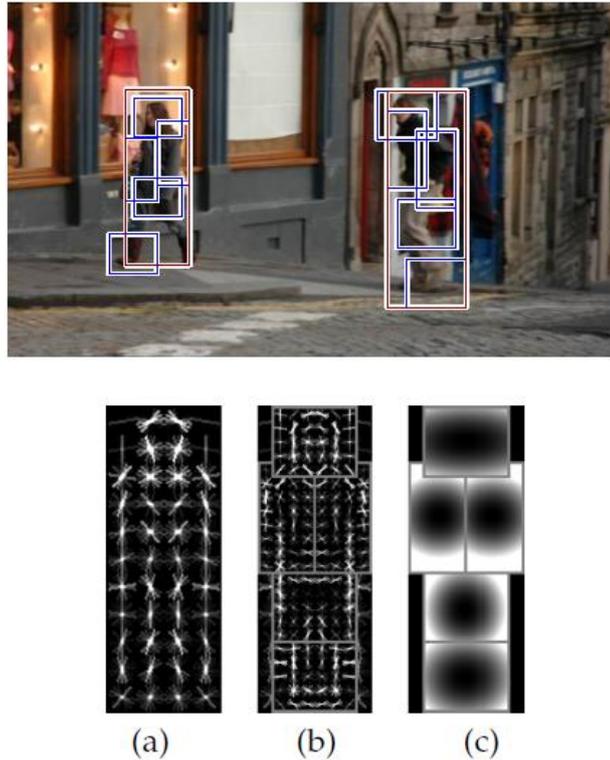
Este capítulo describe en profundidad los algoritmos que han sido seleccionados para llevar a cabo el seguimiento de múltiples personas en vídeo en este proyecto. Para ello, se analiza detalladamente el detector Latent SVM (sección 4.2) y los algoritmos de seguimiento Filtro de Kalman (sección 4.3) y Filtro de Partículas (sección 4.4).

### 4.2 Detector Latent SVM

El detector Latent SVM [33] se basa en la búsqueda exhaustiva y en un modelo de persona basado en partes. Se trata de una adaptación basada en partes del detector HOG original. Propone un sistema de detección de objetos basado en la combinación de modelos deformables de partes en multiescala donde cada parte del cuerpo deformable se modela como el detector HOG original.

El detector de Dalal-Triggs [34] utiliza un solo filtro basado en características de histogramas de gradientes orientados (HOG) para representar a una categoría de objeto (en nuestro caso, personas). Este detector utiliza un enfoque de ventana deslizante que aplica un filtro en cada una de las posiciones y escalas de una imagen. Podemos pensar en el detector como un clasificador que toma como entrada una imagen, una posición dentro de esa imagen, y una escala. El clasificador determina si existe o no una instancia de la categoría del objetivo en la posición y la escala dada. Debido a que el modelo es un filtro simple podemos calcular una puntuación como  $\beta * \Phi(x)$ , donde  $\beta$  es el filtro,  $x$  es una imagen con una posición especificada y escala, y  $\Phi(x)$  es un vector de características. Por el contrario, el detector Latent SVM se basa en un modelo estructurado basado en partes definido por un filtro raíz análogo al filtro de Dalal-Triggs más un conjunto de filtros de partes y modelos de deformación asociados. La puntuación de este modelo en una posición y escala particular de una imagen es la puntuación del filtro raíz en la ubicación dada, más la suma de puntuación de cada parte menos un coste de deformación que mide la desviación de la parte de su localización relativa a la raíz. Los filtros de partes capturan características al doble de resolución espacial que las capturadas mediante el filtro raíz. De este modo, se modela la apariencia visual en múltiples escalas. La Figura 4-1 muestra el modelo de la categoría persona de una sola componente.

Todos los modelos implican filtros lineales que se aplican a mapas densos de características. Un mapa de características es una matriz cuyas entradas son vectores de características d-dimensionales calculados a partir de una cuadrícula densa de localizaciones en una imagen. Intuitivamente cada vector de características describe una parte local de la imagen. Como características se utiliza una variación de las características del HOG.



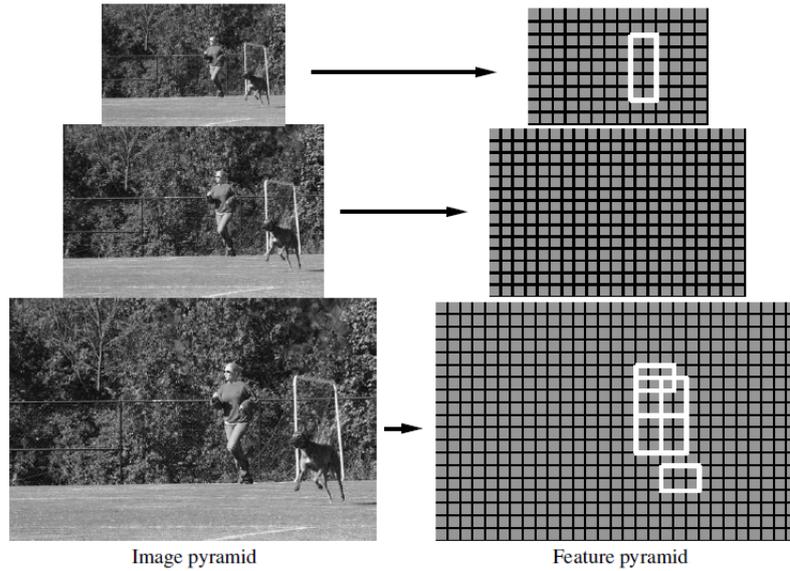
**Figura 4-1: Detecciones obtenidas con un modelo de persona de una sola componente. El modelo se define mediante un filtro raíz (a), varios filtros de partes de mayor resolución (b) y un modelo espacial para la localización de cada parte relativa a la raíz c). Los filtros especifican pesos para las características del histograma de gradientes orientados. Sus visualizaciones muestran los pesos positivos en diferentes orientaciones. La visualización de los modelos espaciales refleja el “coste” de colocar el centro de una parte en diferentes localizaciones relativas a la raíz.**

Un filtro es una plantilla rectangular definida por una matriz de vectores ponderados d-dimensionales. La respuesta, puntuación, de un filtro  $F$  en una posición  $(x, y)$  en un mapa de características  $G$  es el "producto escalar" del filtro y una subventana del mapa de características con la esquina superior izquierda en  $(x, y)$ , es decir:

$$\sum_{x', y'} F[x', y'] \cdot G[x + x', y + y'] \quad \text{Ecuación 19}$$

Para definir una puntuación en diferentes posiciones y escalas de una imagen se utiliza una pirámide de características, que especifica un mapa de características para un número finito de escalas en un rango fijo. En la práctica, se obtienen las pirámides de características calculando una pirámide de imágenes estándar a través de suavizado repetido y submuestreo, y luego calculando un mapa de características de cada nivel de la pirámide de imágenes. La Figura 4-2 ilustra un ejemplo de la obtención de una pirámide de características.

Como se ha comentado anteriormente, los modelos se definen por un filtro raíz que cubre aproximadamente un objeto completo y por una serie de filtros de partes de mayor resolución que cubren partes más pequeñas del objeto. La ubicación del filtro raíz define una ventana de detección (los píxeles que contribuyen a la parte del mapa de características



**Figura 4-2: Una pirámide de características y una instancia de un modelo de persona dentro de dicha pirámide. Los filtros de partes se colocan al doble de resolución de la posición del filtro raíz.**

cubierto por el filtro). Los filtros de partes se colocan en el nivel de características correspondiente al doble de resolución que el nivel de características del filtro raíz.

El uso de características de mayor resolución para definir los filtros de partes es esencial para obtener alto rendimiento de reconocimiento. Con este enfoque los filtros de partes capturan las características de resolución más fina que se localizan con una mayor precisión que las características capturadas por el filtro raíz. Por ejemplo, si se considera el caso de la elaboración de un modelo para el reconocimiento de una cara, el filtro raíz podría capturar los bordes de baja resolución tales como los límites de la cara mientras que los filtros de partes podrían capturar detalles como los ojos, la nariz y la boca.

Un modelo para un objeto con  $n$  partes se define formalmente por  $n + 2$  componentes  $(F_0, P_1, \dots, P_n, b)$  donde  $F_0$  es un filtro raíz,  $P_i$  es un modelo para  $i$ -ésima parte y  $b$  es un término de sesgo. Cada modelo de parte se define por tres componentes  $(F_i, v_i, d_i)$  donde  $F_i$  es un filtro para la parte  $i$ ,  $v_i$  es un vector bidimensional especificando la posición "ancla" para la parte  $i$  con respecto a la posición de la raíz, y  $d_i$  es un vector de cuatro dimensiones que especifica los coeficientes de una función cuadrática que define un coste de deformación para cada posible colocación de la parte relativa a la posición ancla.

Una hipótesis de objeto especifica la ubicación de cada filtro en el modelo en una pirámide de características,  $z = (p_0, \dots, p_n)$ , donde  $p_i = (x_i, y_i, l_i)$  especifica el nivel y la posición del filtro  $i$ -ésimo. La puntuación de una hipótesis viene dada por las puntuaciones de cada filtro en sus respectivos emplazamientos menos un coste de deformación que depende de la posición relativa de cada parte con respecto a la raíz, más el sesgo, es decir:

$$score(p_0, \dots, p_n) = \sum_{i=0}^n F'_i \cdot \phi_d(H, p_i) - \sum_{i=1}^n d_i \cdot \phi_d(dx_i, dy_i) + b \quad \text{Ecuación 20}$$

donde

$$(dx_i, dy_i) = (x_i, y_i) - (2(x_0, y_0) + v_i) \quad \text{Ecuación 21}$$

proporciona el desplazamiento de la parte relativa  $i$ -ésima a su posición ancla  $y$

$$\phi_d(dx_i, dy_i) = (dx, dy, dx^2, dy^2) \quad \text{Ecuación 22}$$

son características de deformación.

Para detectar objetos en una imagen se calcula una puntuación total para cada ubicación del filtro raíz de acuerdo a la mejor colocación posible de las partes, es decir:

$$score(p_0) = \max_{p_1, \dots, p_n} score(p_0, \dots, p_n) \quad \text{Ecuación 23}$$

En primer lugar, se determinan las respuestas de cada uno de los modelos de filtro para cada nivel de la pirámide de características mediante:

$$R_{i,l}(x, y) = F'_i \cdot \phi(H, (x, y, l)) \quad \text{Ecuación 24}$$

Después de calcular las respuestas de los filtros se transforman las respuestas de los filtros de partes para permitir la incertidumbre espacial, es decir:

$$D_{i,l}(x, y) = \max_{dx, dy} (R_{i,l}(x + dx, y + dy) - d_i \cdot \phi_d(dx, dy)) \quad \text{Ecuación 25}$$

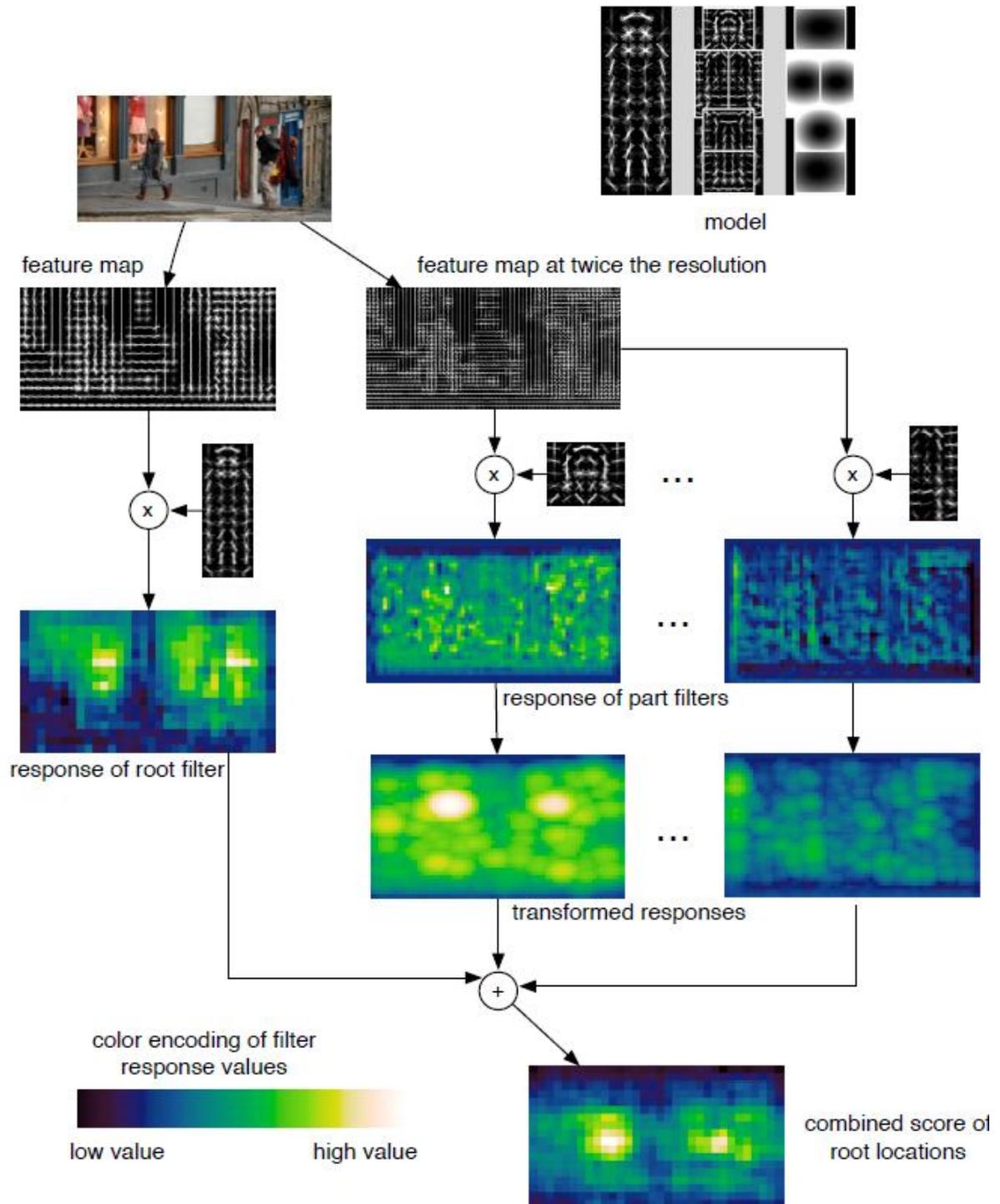
El valor  $D_{i,l}(x, y)$  es la contribución máxima de la parte  $i$ -ésima a la puntuación de una ubicación de la raíz que coloca la posición ancla de esta parte en la posición  $(x, y)$  en el nivel  $l$ . Por tanto, las puntuaciones globales en cada nivel se pueden expresar mediante la suma de la respuesta del filtro raíz en ese nivel más las versiones desplazadas de las respuestas de las partes transformadas y submuestreadas, es decir:

$$score(x_0, y_0, l_0) = R_{0,l_0}(x_0, y_0) + \sum_{i=1}^n D_{i,l_0-\lambda}(2(x_0, y_0) + v_i) + b \quad \text{Ecuación 26}$$

donde  $\lambda$  indica el número de niveles que se necesita bajar en la pirámide de características para obtener un mapa de características que se calcula al doble de resolución que el del filtro raíz.

La Figura 4-3 muestra el procedimiento que sigue el detector Latent SVM para obtener las puntuaciones globales en cada uno de los niveles de la pirámide de características.

Por tanto, a través de este método, se obtiene un mapa de confianza de detección denso. Para obtener detecciones individuales se aplica un algoritmo de *non-maximum-suppression* que elimina detecciones repetidas que se solapan entre ellas.



## 4.3 Filtro de Kalman

### 4.3.1 Algoritmo del Filtro de Kalman

El Filtro de Kalman [52] es un método que genera una única hipótesis de seguimiento en cada *frame*. Consiste en un conjunto de ecuaciones matemáticas que proveen una solución recursiva óptima. Su objetivo principal es determinar un estimador de un sistema en el instante  $t$  a partir de la información disponible en el instante  $t-1$ , y actualizar dicha estimación a partir de la información actual del instante  $t$ .

La solución es óptima ya que el filtro combina toda la información observada y el conocimiento previo acerca del comportamiento del sistema para producir una estimación del estado<sup>1</sup> de tal manera que el error es minimizado estadísticamente. El término recursivo significa que el filtro recalcula la solución cada vez que una nueva observación o medida<sup>2</sup> es incorporada al sistema.

El Filtro de Kalman es el principal algoritmo para estimar sistemas dinámicos representados en la forma de estado-espacio. En esta representación el sistema es descrito por un conjunto de variables denominadas de estado. El estado contiene toda la información relativa al sistema en un cierto punto en el tiempo. Esta información debe permitir la inferencia del comportamiento pasado del sistema, con el objetivo de predecir su comportamiento futuro.

Lo que hace al Filtro de Kalman tan interesante es precisamente su habilidad para predecir el estado de un sistema en el pasado, en el presente y en el futuro, aún cuando la naturaleza precisa del sistema modelado es desconocida. En la práctica, las variables de estado de un sistema dinámico no pueden ser determinadas exactamente por una medición directa por lo que su medición se realiza por medio de procesos estocásticos que involucran algún grado de incertidumbre en la medición.

El proceso seguido por el Filtro de Kalman tiene como objetivo resolver el problema de estimar el estado  $X \in R^n$  de un proceso controlado en tiempo discreto, el cual viene determinado por una ecuación lineal estocástica en diferencias de la siguiente forma:

$$X_t = AX_{t-1} + w_{t-1} \quad \text{Ecuación 27}$$

donde  $X_t$  es el estado en el instante  $t$ ,  $A$  es la matriz de dimensión  $n \times n$  que relaciona el estado en el instante anterior con el estado en el momento actual,  $X_{t-1}$  es el estado en el instante  $t-1$  y  $w_{t-1}$  es el error del proceso en el instante  $t-1$ .

---

<sup>1</sup> El estado contiene la información utilizada para representar el objetivo en cada instante de tiempo (por ejemplo, centroide del *bounding box*, velocidad del objetivo, etc.).

<sup>2</sup> El término medida hace referencia a la información disponible del objetivo presente en el *frame* actual (por ejemplo, *bounding box* del objetivo tras una nueva detección proporcionada por un detector de personas) que se utiliza para corregir el pronóstico del estado.

con una medida  $Z$  correspondiente a la observación y perteneciente a  $R^m$  que viene dada por:

$$\mathbf{Z}_t = \mathbf{H}\mathbf{X}_t + \mathbf{v}_t \quad \text{Ecuación 28}$$

donde  $Z_t$  son las medidas en el instante  $t$ ,  $H$  es la matriz de dimensión  $m \times n$  que relaciona el estado con la medición,  $X_t$  es el estado en el momento  $t$  y  $v_t$  representa el error de la medida en el instante  $t$ .

Las variables  $v_t$  y  $w_t$  que como se ha visto representan el error del proceso y de la medida respectivamente, se asume que son independientes entre ellas, esto es, que son ruido blanco y con distribución de probabilidad normal:

$$p(\mathbf{w}) \cong N(\mathbf{0}, \mathbf{Q}) \quad \text{Ecuación 29}$$

$$p(\mathbf{v}) \cong N(\mathbf{0}, \mathbf{R}) \quad \text{Ecuación 30}$$

donde  $Q$  y  $R$  son las matrices de covarianza de la perturbación del proceso y de la perturbación de la medida, respectivamente.

En la práctica las matrices de covarianza de la perturbación del proceso,  $Q$ , y de la perturbación de la medida,  $R$ , podrían cambiar en el tiempo, aunque por simplicidad en general se asumen constantes.

El Filtro de Kalman estima el proceso anterior utilizando un control de retroalimentación, esto es, estima el proceso en un instante de tiempo y entonces obtiene la retroalimentación por medio de los datos observados. Desde este punto de vista las ecuaciones que se utilizan para derivar el Filtro de Kalman se pueden dividir en dos grupos:

- Ecuaciones que realizan la predicción.
- Ecuaciones que actualizan los datos observados o ecuaciones de actualización.

Las ecuaciones del primer grupo son responsables de la proyección del estado al momento  $t$  tomando como referencia el estado  $t-1$  y de la actualización intermedia de la matriz de covarianza del estado. El segundo grupo de ecuaciones son responsables de la retroalimentación, es decir, incorporan nueva información dentro de la estimación anterior de tal forma que se obtiene una estimación mejorada del estado. Por ello, el algoritmo de Filtro de Kalman se puede considerar como un algoritmo de pronóstico-corrección.

A continuación, se incluyen las ecuaciones específicas de la etapa de predicción:

$$\hat{\mathbf{X}}_t^* = \mathbf{A}\hat{\mathbf{X}}_{t-1} \quad \text{Ecuación 31}$$

donde  $\hat{\mathbf{X}}_t^*$  es el pronóstico del estado en el instante  $t$ ,  $A$  es la matriz de dimensión  $n \times n$  que relaciona el estado anterior con el estado actual y  $\hat{\mathbf{X}}_{t-1}$  es el estado actualizado del instante  $t-1$ .

$$\mathbf{P}_t^* = \mathbf{A}\mathbf{P}_{t-1}\mathbf{A}^T + \mathbf{Q} \quad \text{Ecuación 32}$$

donde  $P_t^*$  es el pronóstico de la covarianza del error en el instante  $t$ ,  $A$  es la matriz de dimensión  $n \times n$  que relaciona el estado anterior con el estado actual,  $P_{t-1}$  es la covarianza del error actualizada en el instante  $t-1$  y  $Q$  es la covarianza de la perturbación aleatoria del proceso que trata de estimar el estado.

A continuación, se incluyen las ecuaciones del proceso de corrección del filtro. La primera tarea durante la corrección de la proyección del estado es el cálculo de la ganancia de Kalman,  $K_t$ . Este factor de ganancia es seleccionado de forma que minimice la covarianza del error de la nueva estimación del estado.

$$K_t = P_t^* H^T (H P_t^* H^T + R)^{-1} \quad \text{Ecuación 33}$$

donde  $K_t$  es la ganancia de Kalman en el instante  $t$ ,  $P_t^*$  es el pronóstico de la covarianza del error en el instante  $t$ ,  $H$  es la matriz de dimensión  $n \times n$  que relaciona el estado con la medición y  $R$  es la covarianza de la perturbación aleatoria de la medida.

El siguiente paso es realmente medir el proceso para obtener y entonces generar una nueva estimación del estado que incorpora la nueva observación.

$$\hat{X}_t = \hat{X}_t^* + K_t (Z_t - H \hat{X}_t^*) \quad \text{Ecuación 34}$$

donde  $\hat{X}_t$  es la actualización de la estimación de estado en el instante  $t$ ,  $\hat{X}_t^*$  es el pronóstico del estado en el instante  $t$ ,  $K_t$  es la ganancia de Kalman en el momento  $t$ ,  $Z_t$  es la medida en el instante  $t$  y  $H$  es la matriz de dimensión  $n \times n$  que relaciona el estado con la medición.

El último paso es obtener una nueva estimación de la covarianza del error.

$$P_t = (I - K_t H) P_t^* \quad \text{Ecuación 35}$$

donde  $P_t$  es la actualización de la estimación de estado en el instante  $t$ ,  $K_t$  es la ganancia de Kalman en el instante  $t$ ,  $H$  es la matriz de dimensión  $n \times n$  que relaciona el estado con la medición y  $P_t^*$  es el pronóstico de la covarianza del error en el instante  $t$ .

Después de las actualizaciones, el proceso es repetido tomando como punto de partida las nuevas estimaciones del estado y de la covarianza del error. En la Figura 4-4 se puede observar el proceso completo del Filtro de Kalman.

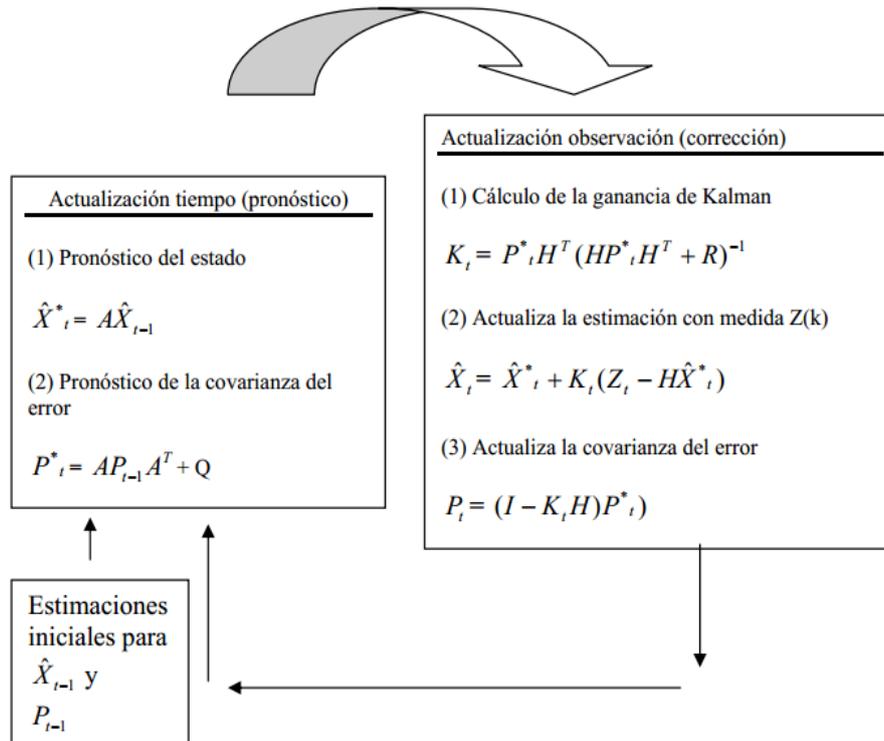


Figura 4-4: Proceso del Filtro de Kalman (extraído de [94])

### 4.3.2 Filtro de Kalman seleccionado

El Filtro de Kalman utilizado en este proyecto se basa en la implementación que incluye la librería de acceso público OpenCV. Se define el estado  $X_t$  mediante los siguientes parámetros:

$$X_t = (x, y, cx, cy, u_x, u_y, u_{cx}, u_{cy}) \quad \text{Ecuación 36}$$

donde  $x$  y  $y$  representan la coordenada superior izquierda del *bounding box* del objetivo,  $cx$  y  $cy$  representan el centroide del *bounding box* del objetivo y  $u_x$ ,  $u_y$ ,  $u_{cx}$  y  $u_{cy}$  representan las componentes de velocidad.

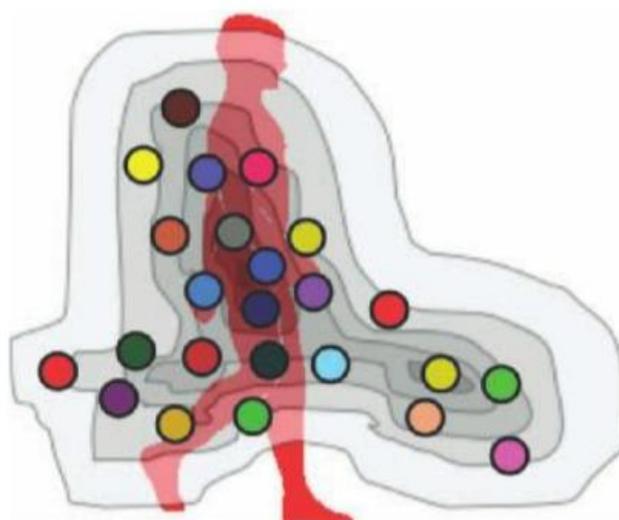
Las matrices de la covarianza de la perturbación del proceso (Q), de la covarianza de la perturbación del proceso (R) y la matriz de la estimación de la covarianza del error en el instante  $t = 0$  ( $P_0$ ) se definen como matrices identidades multiplicadas cada una de ellas por una constante.

En cada iteración (salvo  $t = 0$ ) el Filtro de Kalman lleva a cabo su etapa de predicción en la que proyecta el estado al instante  $t$  tomando como referencia el estado  $t-1$ . Además, por medio de su control de retroalimentación realiza su etapa de corrección para obtener una estimación mejorada del estado, siempre y cuando disponga de la información de medida (en nuestro caso, una nueva detección del objetivo proporcionada por un detector de personas) relativa al objetivo seguido en el instante actual.

## 4.4 Filtro de Partículas

### 4.4.1 Algoritmo del Filtro de Partículas

El Filtro de Partículas [53] es un método que genera múltiples hipótesis de seguimiento en cada *frame* (*Multiple Hypothesis Localisation*). Estas hipótesis se validan en función de las mediciones actuales del *frame* y de un modelo de movimiento establecido del objetivo (*target*). De esta forma, las hipótesis más probables del estado del objetivo se propagan, mientras que las más improbables se eliminan. El uso de múltiples hipótesis permite afrontar mejor problemas como el *clutter*<sup>3</sup> u oclusiones. Sin embargo, son computacionalmente más costosos que los algoritmos de una única hipótesis por lo que imponen reglas más estrictas en cuanto a la dimensionalidad del espacio de estado del objetivo. De hecho, el número de hipótesis que son necesarias para explorar un espacio de estado multidimensional crece exponencialmente con el número de dimensiones en el espacio del estado.



**Figura 4-5: Los métodos de localización de múltiples hipótesis extraen un número variable de hipótesis para después determinarse la calidad de cada muestra (extraído de [1]).**

Este algoritmo se emplea con el fin de estimar el estado de un sistema que varía con el tiempo. Este método, también conocido como SIS (*Sequential Importance Sampling*) o Algoritmo de Condensación, está formado por un conjunto de partículas que representan los estados posibles del objetivo en el espacio y por un conjunto de pesos asociados con cada partícula, mediante los cuales se representan los valores a posteriori de la función de densidad de probabilidad. En la Figura 4-5 se puede observar el conjunto de partículas que representan las múltiples hipótesis del estado del objetivo que se extraen en cada *frame*.

La idea fundamental de este algoritmo es representar la densidad posterior mediante un conjunto de partículas aleatorias con pesos asociados y entonces estimar el cálculo del estado del objetivo basándose en estos datos. Por tanto, su objetivo es determinar el estado

---

<sup>3</sup> El fenómeno de *clutter* aparece cuando la apariencia de los objetivos y del fondo es similar, de tal forma que puede ocasionar errores en el proceso de seguimiento del objetivo que está siendo seguido.

$x_k$  a partir de las observaciones  $z_k$ , siendo el estado óptimo determinado por la siguiente relación:

$$\hat{x}_k = E[x_k | Z_0^k] \quad \text{Ecuación 37}$$

donde  $Z_0^k$  es el conjunto de observaciones<sup>4</sup> hasta el instante  $k$ .

En el proceso de seguimiento se distinguen las siguientes etapas para el Filtro de Partículas:

- 1) Modelo: Se crea un modelo del objeto a seguir a partir de las características extraídas durante el proceso de inicialización, por ejemplo, utilizando histogramas de color.
- 2) Inicialización de las partículas: Para llevar a cabo el seguimiento del objetivo, se crean un conjunto de puntos de forma aleatoria o utilizando algún tipo de información disponible (posición aproximada, tamaño del objetivo, etc.).
- 3) Predicción: Una vez creadas las partículas de los *frames* anteriores, se calculan pequeñas modificaciones en el estado, por ejemplo, la adición de ruido que contribuirá a la variabilidad del sistema. Esto ayudará a estimar el estado del objetivo en el *frame* actual.
- 4) Actualización: A cada muestra se le asigna un peso en función de la similitud con el modelo del objetivo. La información del *frame* actual se utiliza para calcular su similitud.
- 5) Remuestreo: Esta etapa permite eliminar partículas con pesos bajos, descartándose de este modo hipótesis improbables y evitándose que después de algunas iteraciones se propaguen partículas con pesos insignificantes. Este método se conoce generalmente como empobrecimiento de partículas (*particle impoverishment*).
- 6) Actualización del modelo: Este paso es opcional y permite actualizar el modelo del objetivo generado en el proceso de inicialización mediante la información del *frame* actual.

Una vez que se lleva a cabo el paso de remuestreo, el proceso se repite desde la etapa 3 a la 6 hasta que la secuencia termina o se elimina el objetivo.

#### 4.4.2 Filtro de Partículas seleccionado

El Filtro de Partículas seleccionado se basa en la idea de [96] pero con algunas modificaciones. Para cada instante de tiempo, la salida del filtro es el conjunto de  $N$  partículas ponderadas  $X_t = \{x_t^{(n)}, \pi_t^{(n)}\}_{n=1, \dots, N}$ , donde  $x_t^{(n)}$  representa un estado hipotético

---

<sup>4</sup> Las observaciones  $z_k$  representan el conjunto de partículas aleatorias disponibles en el instante  $k$  con pesos asociados en función de la similitud con el modelo del objetivo definido que se utilizan para estimar el estado del objetivo en ese instante.

del objetivo y  $\pi_t^{(n)}$  representa el peso asociado a la partícula  $n$  que depende de la similitud entre el modelo del objetivo del estado actual y el modelo extraído del estado  $x_t^{(n)}$ . Cada estado  $x_t^{(n)}$  en el instante  $t$  para el Filtro de Partículas seleccionado se define mediante los siguientes parámetros:

$$\mathbf{x}_t^{(n)} = (x, y, u_x, u_y) \quad \text{Ecuación 38}$$

donde  $x$  y  $y$  representan la posición del objetivo y  $u_x$  y  $u_y$  representan las componentes de velocidad.

Dado el conjunto de partículas  $X_{t-1}$  y el modelo del objetivo  $q$ , se define el estado actual de cada una de ellas como:

$$\mathbf{x}_t^{(n)} = \mathbf{A}\mathbf{x}_{t-1}^{(n)} + \mathbf{w}_{t-1}^{(n)} \quad \text{Ecuación 39}$$

donde  $x_t^{(n)}$  representa el estado de cada partícula en el instante  $t$ ,  $A$  define la matriz de transición del sistema con un modelo de velocidad constante,  $x_{t-1}^{(n)}$  representa el estado de cada partícula en el instante  $t-1$  y  $w_{t-1}^{(n)}$  es una variable aleatoria Gaussiana multivariable que modela el proceso de ruido.

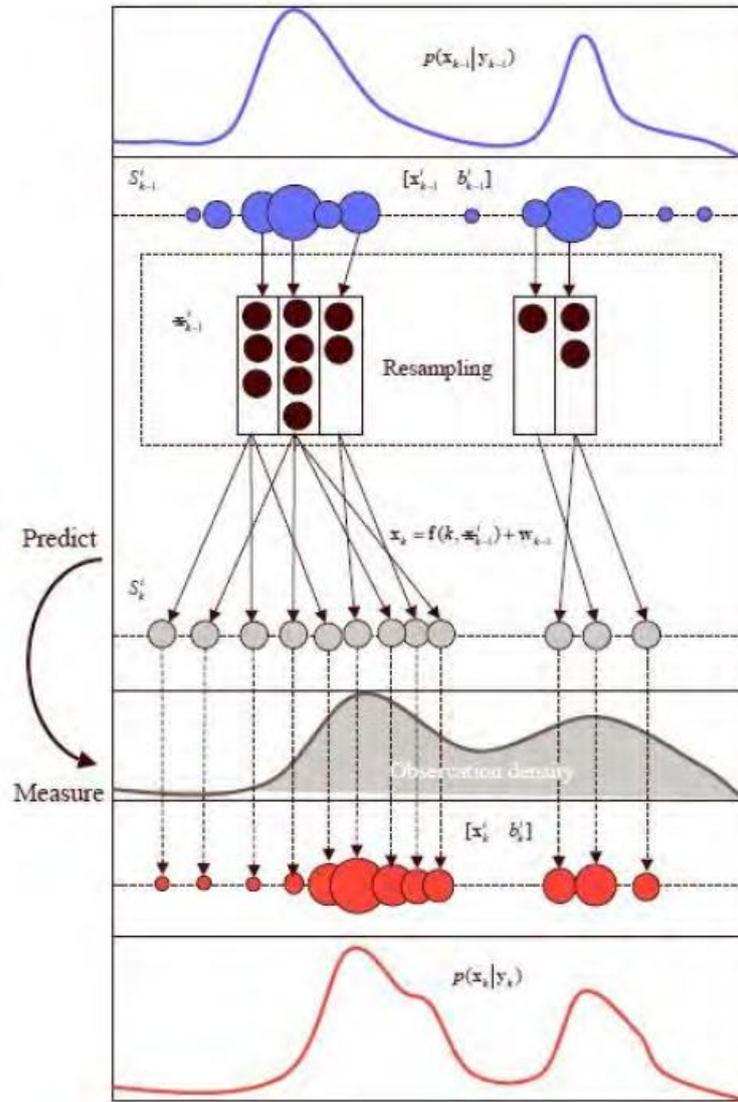


Figura 4-6: Esquema general del proceso de seguimiento de un Filtro de Partículas (extraído de [95]).

A continuación, se describe el proceso llevado a cabo durante una iteración del Filtro de Partículas basado en color:

- 1) Cálculo del histograma de color para cada partícula  $X_t$ :

$$\mathbf{p}_{x_t^{(n)}} = \{\mathbf{p}_{x_t^{(n)}}^{(u)}(x)\}_{u=1}^{N_b} \quad \text{Ecuación 40}$$

donde  $\mathbf{p}_{x_t^{(n)}}$  representa el histograma de color de la partícula  $n$ ,  $N_b$  es el número de cajas del histograma y la probabilidad de cada característica  $\mathbf{p}_{x_t^{(n)}}^{(u)}(x)$  es:

$$\mathbf{p}_{x_t^{(n)}}^{(u)}(x) = \mathbf{C}_{x_t^{(n)}} \sum_{i=1}^M \delta(\mathbf{b}(w_i) - u) \quad \text{Ecuación 41}$$

donde  $C_{x_t^{(n)}}$  es una constante de normalización requerida para asegurar que  $\sum_{u=1}^{N_b} p_{x_t^{(n)}}^{(u)}(x) = 1$ ,  $\delta$  es la delta de Kronecker,  $w_i$  las localizaciones de los píxeles y  $b(w_i)$  es una función que asocia el píxel en la posición  $w_i$  a su correspondiente caja del histograma.

- 2) Medida de similitud entre el histograma calculado para cada partícula  $X_t$  con el histograma de color del modelo mediante el cálculo del coeficiente Bhattacharyya para la obtención del peso de cada partícula mediante la siguiente relación:

$$\pi_t^{(n)} = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(1-\rho[p_{x_t^{(n)}}^{(n)}, q])}{2\sigma^2}} \quad \text{Ecuación 42}$$

donde  $\pi_t^{(n)}$  representa el peso de cada partícula en el instante  $t$ ,  $\sigma$  es la varianza de la Gaussiana y  $\rho[p_{x_t^{(n)}}^{(n)}, q]$  es el coeficiente Bhattacharyya que viene dado por:

$$\rho[p_{x_t^{(n)}}^{(n)}, q] = \sum_{u=1}^{N_b} \sqrt{p_{x_t^{(n)}}^{(u)}(x), q^u} \quad \text{Ecuación 43}$$

donde  $q$  es el modelo del objetivo,  $p_{x_t^{(n)}}^{(n)}$  representa el histograma de color de la partícula  $n$  y  $N_b$  es el número de cajas del histograma.

- 3) Estimación del estado medio del conjunto  $X_t$ :

$$E[X_t] = \sum_{n=1}^N \pi_t^{(n)} x_t^{(n)} \quad \text{Ecuación 44}$$

donde  $\pi_t^{(n)}$  representa el peso de cada partícula en el instante  $t$  y  $x_t^{(n)}$  es el estado de cada partícula en el instante  $t$ .

- 4) Remuestreo sistemático: Se seleccionan  $N$  partículas del conjunto de partículas  $X_t$ . Esta selección se lleva a cabo basándose en el peso de cada una de las partículas. Una partícula se selecciona hasta que su confianza es menor que la confianza acumulativa esperada.

$$X'_t = f(X_t) = f(\{x_t^{(n)}, \pi_t^{(n)}\}_{n=1, \dots, N}) \quad \text{Ecuación 45}$$

$$X'_t = \{x'_t^{(n)}, \pi'_t^{(n)}\}_{n=1, \dots, N} \quad \text{Ecuación 46}$$

donde  $X_t$  es el conjunto de partículas en el instante  $t$ ,  $X'_t$  es el nuevo conjunto de partículas creado tras el proceso de remuestreo,  $x_t^{(n)}$  representa

el estado de cada partícula del conjunto  $X_t$  en el instante  $t$ ,  $\pi_t^{(n)}$  representa el peso de la partícula  $n$  del conjunto  $X_t$  en el instante  $t$ ,  $x_t'^{(n)}$  representa el estado de cada partícula del nuevo conjunto  $X'_t$  en el instante  $t$  y  $\pi_t'^{(n)}$  representa el peso de la partícula  $n$  del nuevo conjunto  $X'_t$  en el instante  $t$ .

- 5) Propagación de cada partícula del nuevo conjunto mediante una ecuación diferencial estocástica:

$$\mathbf{x}_{t+1}^{(n)} = \mathbf{A}\mathbf{x}_t'^{(n)} + \mathbf{w}_t^{(n)} \quad \text{Ecuación 47}$$

donde  $x_{t+1}^{(n)}$  representa el estado de cada partícula en el instante  $t+1$ ,  $\mathbf{A}$  define la matriz de transición del sistema con un modelo de velocidad constante,  $x_t'^{(n)}$  representa el estado del nuevo conjunto de partículas creado tras el proceso de remuestreo y  $w_t^{(n)}$  es una variable aleatoria Gaussiana multivariable que modela el proceso de ruido.

## 4.5 Conclusiones

Durante este capítulo se han detallado los algoritmos que han sido seleccionados para llevar a cabo el algoritmo de seguimiento de múltiples personas en este proyecto. Para ello, se ha analizado detalladamente el detector Latent SVM (sección 4.2) y los algoritmos de seguimiento Filtro de Kalman (sección 4.3) y Filtro de Partículas (sección 4.4).

En la sección 4.2 se ha explicado el procedimiento que sigue el detector Latent SVM. Realiza la detección de los objetos que son candidatos a ser persona por medio de una búsqueda exhaustiva por lo que es robusto a la rotación y a cambios de escala y pose, pero en cambio, su coste computacional es superior. Escanea la imagen de entrada en múltiples escalas y ubicaciones de tal forma que busca similitudes con el modelo de persona elegido. A través de este método, se obtiene un mapa de confianza de detección denso mediante el cual se obtienen detecciones individuales tras aplicar algún tipo de *non-maximum-suppression* que elimina detecciones repetidas que se solapan entre ellas. Además, define el modelo de persona haciendo uso de la información de apariencia. Describe a la persona como combinación de múltiples regiones o formas, es decir, se trata de un modelo basado en partes y define su apariencia de acuerdo a sus características de información de borde utilizando el descriptor de forma HOG. Por tanto, al ser un modelo basado en partes se añade complejidad al algoritmo pero es más robusto a variaciones de la gente y a oclusiones.

En la sección 4.3 se ha explicado que el Filtro de Kalman tiene como objetivo determinar un estimador de un sistema en el instante actual a partir de la información disponible en el instante anterior, y actualizar dicha información a partir de la información del instante actual. La solución es óptima ya que el filtro combina toda la información observada y el conocimiento previo acerca del comportamiento del sistema para producir una estimación del estado (que contiene toda la información relativa al sistema en un cierto punto en el tiempo) de tal manera que el error es minimizado estadísticamente. Por tanto, estima el proceso anterior utilizando un control de retroalimentación, esto es, estima el proceso en un

instante de tiempo y entonces obtiene la retroalimentación por medio de los datos observados. Al tratarse de un método que genera una única hipótesis de seguimiento en cada *frame* su coste computacional es menor que los algoritmos que generan múltiples hipótesis en cada instante de tiempo, sin embargo, debido a esto es menos robusto a oclusiones del objetivo a seguir.

Por último, en la sección 4.4 se ha explicado que el Filtro de Partículas tiene como objetivo estimar el estado de un sistema que varía con el tiempo representando la densidad posterior mediante un conjunto de partículas aleatorias con pesos asociados y entonces estimar el cálculo del estado del objetivo basándose en estos datos. Por tanto, se trata de un método que genera múltiples hipótesis de seguimiento en cada *frame* que se validan en función de las mediciones actuales del *frame* y de un modelo de movimiento establecido del objetivo. El uso de múltiples hipótesis permite que el algoritmo afronte mejor problemas como las oclusiones, pero es computacionalmente más costoso que un algoritmo que genera una única hipótesis en cada instante de tiempo.

# 5 Sistema base y módulos comunes

## 5.1 Introducción

En este capítulo se presenta el sistema base que se ha desarrollado para el seguimiento de múltiples personas basado en detección. En primer lugar, se indica la arquitectura general del sistema (sección 5.2). Seguidamente, se describe cada uno de los módulos que conforman dicha arquitectura general y sus características (sección 5.3). Posteriormente, se incluye un apartado que contiene los valores empíricos de cada uno de los parámetros que contiene el sistema base (sección 5.4). Y por último, se incluyen los resultados obtenidos para dicho sistema (sección 5.5) y una serie de conclusiones extraídas de cada una de las secciones que contiene el capítulo (sección 5.6).

## 5.2 Arquitectura del sistema

El objetivo principal en este trabajo ha sido crear un *framework* de seguimiento de personas basado en la combinación de los resultados proporcionados por un detector de personas y por un *tracker*. El diagrama de bloques de la arquitectura del sistema propuesta se muestra en la Figura 5-1. Como punto de partida, se crea un sistema base sobre el cual aportar mejoras en el módulo de *Matching*.

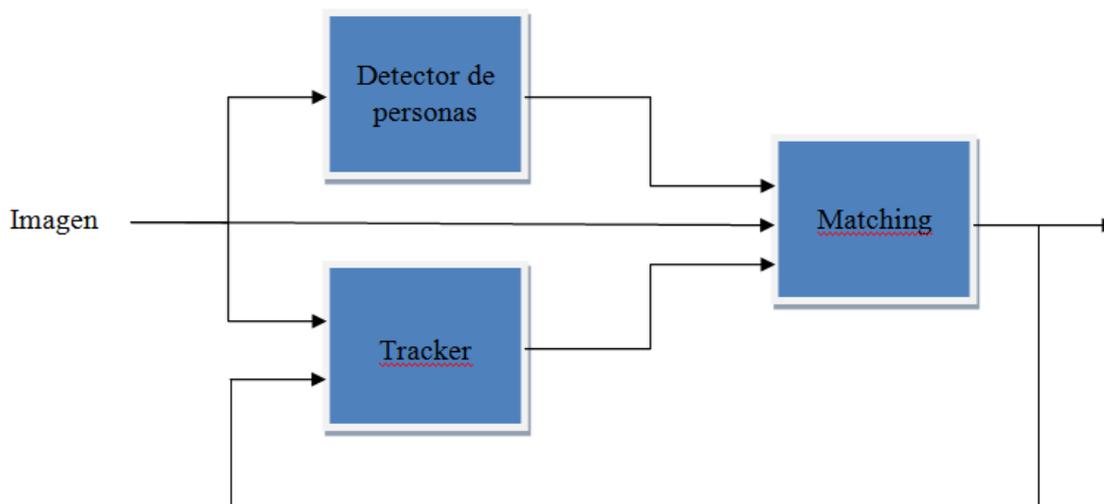


Figura 5-1: Diagrama de bloques del sistema.

## 5.3 Sistema base

En esta sección se describe el sistema base realizado en este proyecto para el seguimiento de personas en vídeo.

### 5.3.1 Módulo Detector de Personas

Después de la adquisición del *frame* actual de la secuencia, se aplica un detector de personas que genera como salida *bounding boxes* en aquellas áreas de la escena donde el detector toma la decisión de clasificar dicho área como persona. Para ello, recibe como entrada la imagen actual que se está procesando para obtener, en primer lugar, los *blobs* candidatos de la escena a ser una persona (utilizando alguna técnica de segmentación o de búsqueda exhaustiva). Después, lleva a cabo una función de verificación y clasificación en la que el detector aplica un modelo de persona definido o previamente entrenado a los objetos candidatos a ser una persona. De acuerdo con la comparación o similitud calculada en la etapa de verificación el detector toma una decisión que puede ser binaria (persona o no persona) o difusa (un valor de confianza o probabilidad de ser una persona).

En la sección 4.2 se ha explicado detalladamente el procedimiento de detección que seguía el detector Latent SVM utilizado en este proyecto.

### 5.3.1 Módulo *Tracker*

En paralelo a la ejecución del detector de personas del sistema, en el módulo *Tracker* se realiza la predicción de la nueva posición de cada uno de los objetivos en la escena a partir de la información obtenida en los *frames* anteriores, salvo en el proceso de inicialización de un nuevo objetivo en el que se crea una instancia del *tracker* haciendo uso de la información proporcionada del detector (*bounding box* del objetivo).

En la sección 4.3 y 4.4 se ha explicado más detalladamente el procedimiento que seguía cada uno de los *trackers* utilizados en el proyecto en los procesos de inicialización y predicción.

### 5.3.2 Módulo de *matching* básico

Una vez obtenidas las detecciones del *frame* actual proporcionadas por el detector del sistema base y realizadas las predicciones de los objetivos presentes en la escena por medio del *tracker* utilizado, en el módulo de *Matching* del sistema base se lleva a cabo las asociaciones entre las detecciones del *frame* actual y las detecciones de *frames* anteriores asignando un único ID a cada objetivo, el nacimiento de nuevos objetivos que aparecen en la escena y la muerte de objetivos que llevan un tiempo sin tener ninguna asociación entre detecciones de *frames* anteriores.

#### 5.3.2.1 Inicialización de objetivos

El problema del seguimiento necesita de un proceso de inicialización de los objetivos que van a ser seguidos a lo largo de la escena. La utilización de detectores que localicen dichos objetivos en la escena proporciona inicialización automática en el sistema.

En nuestro sistema, se aplica un detector de personas en cada *frame* de la secuencia. Las salidas del detector que no se asocian con detecciones de *frames* anteriores, se inicializan

como un nuevo objetivo asignándole un ID único y característico. Además, si un nuevo objetivo es creado, se inicializa también una instancia del *tracker* utilizado en el sistema haciendo uso de la información proporcionada de la medida del detector (*bounding box* del objetivo).

Uno de los problemas de la utilización de los detectores de personas como proceso de inicialización automática, es que su salida no es completamente fiable y puede generar falsos positivos que degraden notablemente los resultados objetivos del proceso de seguimiento. Para eliminar gran parte de los falsos positivos que genera un detector de personas se ha incluido un parámetro en el sistema denominado MIN\_ACTIVE mediante el cual el sistema final no obtiene salida del *bounding box* del objetivo hasta que al menos se hayan realizado el número de asociaciones indicado por este parámetro entre detecciones de *frames* consecutivos.

### 5.3.2.2 Asociación de objetivos

El seguimiento múltiple necesita de un proceso de resolución del problema de asociación de identidades de cada objetivo entre *frames* consecutivos. El sistema se ha diseñado de tal forma que una detección se asigna como máximo a un único objetivo, de tal forma que se evita que en la secuencia aparezcan varios objetivos con un mismo identificador. De esta forma, se evitan intercambios de identidades que se pueden ocasionar por errores del detector de personas durante el proceso de seguimiento de un objetivo y pérdidas del *tracker* del objetivo ocasionadas por una incorrecta actualización de su modelo.

Para llevar a cabo el proceso de asociación de identidades se han definido cuatro umbrales: RADIO, THRESHOLD\_H, THRESHOLD\_W y THRESHOLD\_FEATURE. A continuación, se detalla el significado de cada uno de ellos:

- RADIO: Especifica la distancia máxima que puede existir entre el centroide de una nueva detección en el *frame* actual y el centroide de la predicción realizada del *tracker* de un objetivo.
- THRESHOLD\_H: Indica la diferencia máxima en la altura del *bounding box* de una nueva detección en el *frame* actual y del *bounding box* de la predicción realizada por el *tracker* de un objetivo.
- THRESHOLD\_W: Indica la diferencia máxima en el ancho del *bounding box* de una nueva detección en el *frame* actual y el *bounding box* de la predicción realizada por el *tracker* de un objetivo.
- THRESHOLD\_FEATURE: Indica la diferencia máxima entre la característica elegida a medir (por ejemplo, media de color, histogramas de color, descriptores SIFT, etc.) del *bounding box* de una nueva detección en el *frame* actual y del *bounding box* de la predicción realizada por el *tracker* de un objetivo.

Para determinar la distancia entre el centroide de una nueva detección en el *frame* actual y el centroide de la predicción realizada del *tracker* de un objetivo, se ha utilizado la distancia Euclídea frente a otras del estado del arte por su sencillez, su bajo coste computacional y por los buenos resultados que se obtienen con ella:

$$d_{centroide}(det, tr) = \sqrt{(x_{det} - x_{tr})^2 + (y_{det} - y_{tr})^2} \quad \text{Ecuación 48}$$

Para determinar la diferencia entre la característica elegida a medir (por ejemplo, media de color, histogramas de color, SIFT, etc.) del *bounding box* de una nueva detección en el *frame* actual y del *bounding box* de la predicción realizada por el *tracker* de un objetivo se ha utilizado la distancia Euclídea frente a otras del estado del arte por su sencillez, su bajo coste computacional y por los buenos resultados que se obtienen con ella:

$$d_{feature}(det, tr) = \sqrt{\sum_i (f_{det}^i - f_{tr}^i)^2} \quad \text{Ecuación 49}$$

Por tanto, si una nueva detección en el *frame* actual no supera los umbrales descritos anteriormente con una predicción realizada del *tracker* de un objetivo existente en el sistema, dicha detección se asocia a este objetivo.

El *bounding box* de un objetivo con una nueva detección asociada a la salida final del sistema viene dado por la siguiente relación:

$$BB_{final} = (1 - \alpha) * BB_{det} + \alpha * BB_{tr} \quad \text{Ecuación 50}$$

donde  $BB_{final}$  es el *bounding box* del objetivo a la salida del sistema de seguimiento,  $BB_{det}$  es el *bounding box* de la nueva detección asociada y  $BB_{tr}$  es el *bounding box* de la predicción realizada por el *tracker*.

Además, el nuevo *bounding box* obtenido de la combinación de la nueva detección asociada y de la predicción realizada por el *tracker* se utiliza también para corregir el modelo del *tracker* utilizado en el sistema de seguimiento.

Por último, en el sistema pueden existir objetivos activos sin ninguna detección asociada. Esto puede ser debido a que el detector no haya obtenido salida para dichos objetivos (*missing detections*) o porque dichos objetivos se encuentren ocluidos. Se ha definido un parámetro MAX\_NO\_DETECTION mediante el cual se puede utilizar la información del *tracker* para mantener el objetivo activo y obtener salida en el sistema. Este parámetro indica el número de *frames* de predicciones válidas para obtener salida en el sistema para un objetivo de la escena sin tener detección asociada. Tanto en el caso de *missing detections* como en el caso de que los objetivos se encuentren ocluidos, el histórico de la trayectoria que sigue el objetivo continúa siendo el mismo que el obtenido haciendo uso de la información del *tracker*, por lo que al salir de la oclusión o al volver a obtener detección asociada en el histórico no se realiza ningún tipo de interpolación. El sistema se ha diseñado de esta forma debido a que se hace frente a oclusiones de corta duración (siempre que el *tracker* sea capaz de afrontarlas y de alrededor de un segundo de duración) en las que el *tracker* prácticamente no pierde precisión en el seguimiento del objetivo y por tanto, no produce falsos positivos que deterioren los resultados objetivos del sistema de seguimiento de múltiples personas.

### 5.3.2.2.1 Asociación de objetivos mediante algoritmo tipo Greedy

Se ha incorporado un algoritmo tipo *Greedy* para realizar la asociación de identidades del sistema entre detecciones de *frames* consecutivos en el módulo de *matching* del sistema base. Como se ha explicado anteriormente, se han utilizado cuatro umbrales para realizar la asociación de objetivos (RADIO, THRESHOLD\_H, THRESHOLD\_W y THRESHOLD\_FEATURE). Sin embargo, la inclusión de este algoritmo mejora los resultados del seguimiento ya que la asociación por distancia es de esta forma mucho más robusta que la del sistema actual.

Para la introducción del algoritmo de tipo *Greedy* en el sistema se ha seguido el siguiente procedimiento:

- 1) Se crea una matriz de dimensiones MxN, donde M es la dimensión del conjunto de todos los *trackers* del sistema y N es la dimensión del conjunto de todas las nuevas detecciones obtenidas en el *frame* actual.
- 2) Se determinan las distancias para cada par *tracker*-detección utilizando la distancia Euclídea frente a otras del estado del arte por su sencillez, su bajo coste computacional y por los buenos resultados que se obtienen con ella:

$$d(det, tr) = \sqrt{(x_{det} - x_{tr})^2 + (y_{det} - y_{tr})^2} \quad \text{Ecuación 51}$$

- 3) Se selecciona iterativamente el par *tracker*-detección de distancia mínima de la matriz MxN hasta que el valor mínimo seleccionado de la matriz sea igual a infinito.
- 4) Se comprueba si el par *tracker*-detección seleccionado supera los cuatro umbrales definidos del sistema. Si esto ocurre se asocia dicha detección a dicho *tracker*.
- 5) El valor seleccionado de la matriz MxN se inicializa a infinito para no volver a seleccionar este valor en la próxima iteración.
- 6) Para evitar que en el sistema aparezcan distintos objetivos con un mismo identificador se permite como máximo asociar una detección a un objetivo. Por ello, si en la iteración existe asociación entre un *tracker* y una nueva detección se coloca a infinito toda la fila del *tracker* de la matriz de asociación, así como, toda la columna del detector de dicha matriz.

La inclusión de este algoritmo incrementa el coste computacional del sistema pero soluciona problemas de intercambio de identidades que se producían entre objetivos que se encontraban cercanos uno del otro mejorando de esta forma el comportamiento y los resultados objetivos del seguimiento de personas.

---

**Algorithm** Greedy data association

---

$T$  : conjunto de todos los trackers  
 $D$  : conjunto de todas las detecciones  
 $S(tr,d)$  : distancias para cada par tracker-detección  
 $A(tr,d) = 0$  : asociaciones finales de detección  $d$  y tracker  $tr$

```
while no se hayan calculado todas las distancias para cada par tracker-detección
     $S(tr, d) = \sqrt{(x_{det} - x_{tr})^2 + (y_{det} - y_{tr})^2}$ 
end
while  $\min(S(tr,d)) \neq \infty$ 
     $(tr^*, d^*) = \arg \min(S(tr, d))$ 
    if  $S(tr^*, d^*) \geq \tau$  then
         $A(tr,d) = 1$ 
    end
     $S(tr,d) = \infty$ 
    if  $A(tr,d) = 1$  then
        for  $(i=0; i < \text{length}(T); i++)$ 
             $S(tr_i, d) = \infty$ 
        end
        for  $(i=0; i < \text{length}(D); i++)$ 
             $S(tr, d_i) = \infty$ 
        end
    end
end
```

**Figura 5-2:** Algoritmo de asociación tipo *Greedy*.

### 5.3.2.3 Terminación de objetivos

El problema de seguimiento necesita también de un proceso de terminación o muerte de los objetivos en la escena, por ejemplo, cuando salen de la escena o de la visión de la cámara.

En la arquitectura desarrollada se ha afrontado este problema mediante la creación de un parámetro MAX\_INACTIVE mediante el cual se elimina un objetivo inactivo de la escena cuando tras el número de *frames* indicados por este parámetro el objetivo no ha realizado ninguna asociación con detecciones de *frames* posteriores. Un objetivo se considera que se encuentra inactivo en el sistema si como se explicaba en la sección 5.3.2.1 no ha superado el número de asociaciones indicado por el parámetro MIN\_ACTIVE entre detecciones de *frames* consecutivos, y por tanto, aún no ha sido inicializado, o si un objetivo activo en el sistema supera el número de predicciones permitidas sin detección asociada del *tracker* definido por el parámetro MAX\_NO\_DETECTION. Un objetivo que no se encuentra activo en el sistema no genera salida visual, y por tanto, no se guardan sus datos generados en el proceso de seguimiento. Por tanto, este parámetro aparte de eliminar objetivos que salen de la escena también nos permite eliminar falsos positivos generados por el detector de personas utilizado.

### 5.3.2.4 Formato de salida y visualización del sistema

La salida final del sistema está formada por el *bounding box* de cada uno de los objetivos activos en la secuencia mediante la relación descrita en la sección 5.3.2.2 y su identificador característico correspondiente. Los datos característicos de dichos objetivos se almacenan

en cada iteración en un archivo de resultados con formato de salida igual al que se ha descrito en la sección 3.2.1.

Además, la salida visual del sistema para cada objetivo se corresponde de un *bounding box* que delimita el área cubierto de dicho objetivo, de su ID característico y de la trayectoria que sigue durante el transcurso de la escena (ver sección 3.3).

## 5.4 Ajuste de parámetros del sistema base a las características de los vídeos

En esta sección se indican los valores de cada uno de los parámetros existentes en los módulos de la arquitectura propuesta para este proyecto.

### 5.4.1 Módulo Detector de Personas

#### 5.4.1.1 Detector Latent SVM

Los parámetros ajustables para el detector Latent SVM son el umbral de confianza de la detección y el umbral de solapamiento del algoritmo de *non-maximum-suppression*. Tras diversas pruebas experimentales se ha decidido fijar el primer umbral a 0.098 para los primeros 5 niveles del *dataset* y a 0.298 para el último nivel y el segundo umbral a 0.2 para todos los niveles. En el último nivel se ha decidido aumentar un poco el umbral del detector debido a que realiza búsquedas de personas en escalas más pequeñas de la imagen lo que aumenta la aparición de falsos positivos. En la Figura 5-3 se puede observar la importancia de una buena elección del umbral de confianza de la detección para evitar la aparición de un gran número de falsos positivos que degraden notablemente los resultados del sistema de seguimiento de múltiples personas.



Figura 5-3: Ejemplo de una imagen de la salida del detector de personas Latent SVM para un umbral de confianza de la detección de -0.598 (izquierda) y 0.098 (derecha).

## 5.4.2 Módulo Tracker

### 5.4.2.1 Filtro de Kalman

Los parámetros ajustables para el *tracker* Filtro de Kalman seleccionado (ver sección 4.3.2) son las constantes que multiplican a las matrices identidad de la covarianza de la perturbación del proceso (Q) y de la covarianza de la perturbación del proceso (R) y de la matriz identidad de la estimación de la covarianza del error (P) en cada instante de tiempo. Tras diversas pruebas experimentales se ha decidido dejar estos parámetros fijos a los valores que se recomendaba en la implementación utilizada ya que se obtenían resultados muy similares. Estos valores son:  $10^{-5}$  para la constante que multiplica a la matriz identidad Q, de  $10^{-1}$  para la constante que multiplica a la matriz identidad R y de 1 para la constante que multiplica a la matriz identidad P.

### 5.4.2.2 Filtro de Partículas basado en color

Los parámetros ajustables para el *tracker* Filtro de Partículas basado en color seleccionado (ver sección 4.4.2) son  $N$  (número de partículas del filtro),  $\sigma_x$  y  $\sigma_y$  (varianzas de la posición del centroide del objetivo  $x, y$ ) y  $\sigma_{ux}$  y  $\sigma_{uy}$  (varianzas de las dos componentes de velocidad del objetivo). Todos estos parámetros se han obtenido experimentalmente y se ha decidido fijar dichos valores a:  $N = 100$ ,  $\sigma_x = 2$ ,  $\sigma_y = 2$ ,  $\sigma_{ux} = 0.5$  y  $\sigma_{uy} = 0.5$ .

## 5.4.3 Módulo de *Matching* básico

Como se ha explicado en la sección 5.3.2.2, para llevar a cabo el proceso de asociación de identidades se ha utilizado un umbral `THRESHOLD_FEATURE` que indicaba la diferencia máxima entre la característica elegida a medir (por ejemplo, media de color, histogramas de color, descriptores SIFT, etc.) del *bounding box* de una nueva detección en el *frame* actual y del *bounding box* de la predicción realizada por el *tracker* de un objetivo. También, se ha explicado que para determinar la diferencia entre la característica elegida a medir del *bounding box* de una nueva detección en el *frame* actual y del *bounding box* de la predicción realizada por el *tracker* de un objetivo se ha utilizado la distancia Euclídea (véase Ecuación 49).

En nuestro caso, se ha utilizado la media de color como característica representativa por su bajo coste computacional. La diferencia entre la media de color del *bounding box* de una nueva detección en el *frame* actual y del *bounding box* de la predicción realizada por el *tracker* de un objetivo se determina, por tanto, mediante la siguiente relación:

$$d_{color}(det, tr) = \sqrt{(R_{det} - R_{tr})^2 + (G_{det} - G_{tr})^2 + (B_{det} - B_{tr})^2} \quad \text{Ecuación 52}$$

donde R, G y B indican el valor de cada uno de los canales del espacio de color RGB.

Por otro lado, después de realizar diversas pruebas y variaciones de cada uno de los parámetros del módulo de *Matching* básico, se han fijado los siguientes valores:

- $MIN\_ACTIVE = 7$
- $MAX\_INACTIVE = 5$
- $MAX\_NO\_DETECTION = 10$
- $ALPHA = 0.5$

Para los umbrales  $RADIO$ ,  $THRESHOLD\_FEATURE$ ,  $THRESHOLD\_W$  y  $THRESHOLD\_H$  se han fijado unos parámetros específicos para cada secuencia en función de sus características. Estos parámetros se indican en el Anexo A.

## **5.5 Resultados obtenidos del sistema base**

En este apartado se presentan los resultados obtenidos del sistema base implementado para el seguimiento de múltiples personas utilizando el detector Latent SVM y los *trackers* Filtro de Kalman y Filtro de Partículas.

### **5.5.1 Resultados obtenidos utilizando el detector Latent SVM y el *tracker* Filtro de Kalman**

La Tabla 5-1 muestra los resultados medios obtenidos de las secuencias incluidas en cada uno de los niveles del *dataset* (explicado en la sección 3.4) para el detector Latent SVM y el *tracker* Filtro de Kalman. Los resultados obtenidos para cada una de las secuencias del *dataset*, así como los parámetros utilizados en cada una de ellas, se han incluido en el Anexo A (subapartado A.1). Como se puede apreciar, los resultados tanto de MOTA como de MOTP descienden a medida que aumenta la complejidad con cada nivel, a excepción de los resultados obtenidos en MOTA para el nivel 2 y el nivel 3. Esto se debe, a que en el sistema base se está utilizando un detector de personas por partes, que es robusto a oclusiones parciales, por lo que los resultados en el nivel 3 pueden ser más o menos parecidos a los del nivel 2 ya que el número de personas que aparecen en las secuencias es prácticamente el mismo. Por otro lado, se puede observar que el descenso de los resultados es mucho más abrupto para MOTA que para MOTP. Esto se debe a que MOTP medía la precisión entre los objetivos de salida del sistema base y las anotaciones del *ground-truth* para las asociaciones consideradas que superaban un solapamiento mayor del 50%. Por el contrario, MOTA desciende considerablemente porque el número de objetivos en la secuencias incrementan dando lugar a un mayor número de falsos positivos, falsos negativos y cambios de identificador.

Métrica	Resultados Nivel 1	Resultados Nivel 2	Resultados Nivel 3	Resultados Nivel 4	Resultados Nivel 5	Resultados Nivel 6
MOTP	0.73578	0.73032	0.71326	0.70342	0.69903	0.68483
MOTA	0.78848	0.61482	0.67477	0.52726	0.32665	0.08979
FP	0.11054	0.13334	0.03293	0.21177	0.21743	0.29227
TP	0.89902	0.74819	0.70770	0.73902	0.54408	0.38202
FN	0.10099	0.25181	0.29093	0.24625	0.44737	0.59957
ID switch	0.00	0.00	0.80	2.67	3.00	11.33

**Tabla 5-1: Resultados medios obtenidos para cada nivel del *dataset* utilizando el detector Latent SVM y el *tracker* Filtro de Kalman.**

### 5.5.2 Resultados obtenidos utilizando el detector latent SVM y el *tracker* Filtro de Partículas basado en color

La Tabla 5-2 muestra los resultados medios obtenidos de las secuencias incluidas en cada uno de los niveles del *dataset* (explicado en la sección 3.4) para el detector Latent SVM y el *tracker* Filtro de Partículas. Los resultados obtenidos para cada una de las secuencias del *dataset*, así como los parámetros utilizados en cada una de ellas, se han incluido en el Anexo A (subapartado A.2). Como se puede apreciar, el comportamiento es el mismo que el comentado en la sección 5.5.1 con un descenso de los resultados al incrementarse la complejidad de los niveles del *dataset*. Además, se puede observar también, el descenso más abrupto de los resultados para MOTA que para MOTP por el mismo motivo que se comentó en el punto anterior.

Métrica	Resultados Nivel 1	Resultados Nivel 2	Resultados Nivel 3	Resultados Nivel 4	Resultados Nivel 5	Resultados Nivel 6
MOTP	0.72862	0.72540	0.70517	0.69751	0.66426	0.68225
MOTA	0.83606	0.65629	0.64020	0.47343	0.20504	0.23450
FP	0.09911	0.11495	0.04334	0.22339	0.27978	0.21602
TP	0.93517	0.77138	0.68357	0.69681	0.48481	0.45052
FN	0.06483	0.22813	0.31578	0.29414	0.51152	0.53416
ID switch	0.00	0.33	0.60	2.33	2.00	9.33

**Tabla 5-2: Resultados medios obtenidos para cada nivel del *dataset* utilizando el detector Latent SVM y el *tracker* Filtro de Partículas.**

### 5.5.3 Comparativa de resultados obtenidos

La Tabla 5-3 muestra una comparativa de los resultados medios obtenidos para cada uno de los niveles del *dataset* para el detector Latent SVM y los *trackers* Filtro de Kalman y Filtro de Partículas. En negrita se ha resaltado la combinación detector-*tracker* que obtiene mejores resultados. Como se puede apreciar, los resultados para MOTP son prácticamente los mismos para las dos combinaciones de detector-*tracker*, con diferencias de no más de un 1%, a excepción del nivel 5 donde el Filtro de Kalman supera al Filtro de Partículas en un 3%. Esto era de esperar ya que la precisión del sistema de seguimiento de múltiples personas depende en gran medida de la precisión del *bounding box* de la salida del detector de personas utilizado, que en las dos combinaciones es el mismo. Por el contrario, si observamos los resultados obtenidos para MOTA se observa que en unas situaciones una combinación funciona mejor que la otra y en otras sucede lo contrario. Esto puede deberse

a que en determinadas secuencias un *tracker* funcione mejor que el otro produciendo menos pérdidas de los objetivos dando lugar a un mayor número de *true positives* y menos falsos positivos. Sin embargo, se puede apreciar que en el último nivel del *dataset* propuesto el Filtro de Partículas obtiene bastantes mejores resultados que el Filtro de Kalman. Esto se debe a que en este nivel se producen un gran número de oclusiones totales que provocan que el Filtro de Kalman pierda más fácilmente a los objetivos si no siguen una trayectoria lineal, lo que genera un mayor número de falsos positivos y un menor número de *true positives*.

Por lo tanto, se ha tomado la decisión de utilizar la combinación del detector de personas Latent SVM y el *tracker* Filtro de Partículas para añadir las mejoras que se han introducido al sistema base y que se van a explicar en el Capítulo 6, debido a que en el último nivel se ha observado un deterioro considerable en los resultados obtenidos para MOTA con el Filtro de Kalman.

Métrica	Nivel 1			Nivel 2			Nivel 3		
	KALMAN	PARTICLE	Diferencia resultados	KALMAN	PARTICLE	Diferencia resultados	KALMAN	PARTICLE	Diferencia resultados
MOTP	<b>0.73578</b>	0.72862	-0.72%	<b>0.73032</b>	0.72540	-0.49%	<b>0.71326</b>	0.70517	0.81%
MOTA	0.78848	<b>0.83606</b>	4.76%	0.61482	<b>0.65629</b>	4.15%	<b>0.67477</b>	0.64020	3.46%
FP	0.11054	0.09911	1.14%	0.13334	0.11495	1.84%	0.03293	0.04334	1.04%
TP	0.89902	0.93517	3.62%	0.74819	0.77138	2.32%	0.70770	0.68357	2.41%
FN	0.10099	0.06483	3.62%	0.25181	0.22813	2.37%	0.29093	0.31578	2.49%
ID switch	0.00	0.00	0.00	0.00	0.33	-0.33	0.80	0.60	-0.20

Métrica	Nivel 4			Nivel 5			Nivel 6		
	KALMAN	PARTICLE	Diferencia resultados	KALMAN	PARTICLE	Diferencia resultados	KALMAN	PARTICLE	Diferencia resultados
MOTP	<b>0.70342</b>	0.69751	0.59%	<b>0.69903</b>	0.66426	3.48%	<b>0.68483</b>	0.68225	0.26%
MOTA	<b>0.52726</b>	0.47343	5.38%	<b>0.32665</b>	0.20504	12.16%	0.08979	<b>0.23450</b>	14.47%
FP	0.21177	0.22339	1.16%	0.21743	0.27978	6.24%	0.29227	0.21602	7.63%
TP	0.73902	0.69681	4.22%	0.54408	0.48481	5.93%	0.38202	0.45052	6.85%
FN	0.24625	0.29414	4.79%	0.44737	0.51152	6.42%	0.59957	0.53416	6.54%
ID switch	2.67	2.33	-0.34	3.00	2.00	-1.00	11.33	9.33	2.00

**Tabla 5-3: Comparativa de resultados medios obtenidos para cada nivel del *dataset* utilizando el detector Latent SVM y los *trackers* Filtro de Kalman y Filtro de Partículas. La primera columna de cada nivel muestra los resultados obtenidos utilizando el *tracker* Filtro de Kalman, la segunda columna indica los resultados obtenidos utilizando el *tracker* Filtro de Partículas y la tercera columna indica la diferencia en porcentaje entre la mejor y peor combinación de detector-*tracker* en cuanto a los resultados obtenidos para MOTA.**

## 5.6 Conclusiones

Durante este capítulo se ha presentado el sistema base que se ha desarrollado para el seguimiento de múltiples personas basado en detección. Para ello, se ha indicado la arquitectura general del sistema (sección 5.2) y se ha descrito cada uno de los módulos que conforman dicha arquitectura general y sus características (sección 5.3). Además, se ha

incluido un apartado que contiene los valores empíricos determinados de cada uno de los parámetros incluidos en los módulos del sistema base (sección 5.4), así como, se han incluido los resultados objetivos de dicho sistema para cada una de las configuraciones detector-*tracker* utilizadas en este proyecto (sección 5.5).

En la sección 5.2 se ha presentado la arquitectura general del sistema desarrollado para el seguimiento de múltiples personas. El objetivo principal ha sido crear un *framework* de seguimiento de personas basado en la combinación de los resultados proporcionados por un detector de personas y por un *tracker*.

En la sección 5.3 se ha descrito cada uno de los módulos que conforman la arquitectura general del sistema. El Módulo Detector de Personas generaba como salida *bounding boxes* en aquellas áreas de la escena donde el detector tomaba la decisión de clasificar dicho área como persona y el Módulo *Tracker* realizaba la predicción de la nueva posición de cada uno de los objetivos en la escena en el instante actual a partir de la información obtenida en los *frames* anteriores. Una vez se obtenían las detecciones del *frame* actual proporcionadas por el detector y el *tracker* realizaba las predicciones de los objetivos presentes en la escena, en el Módulo de Matching básico se llevaba a cabo las asociaciones entre las detecciones del *frame* actual y las detecciones de *frames* anteriores, el nacimiento de nuevos objetivos que aparecen en la escena y la muerte de objetivos que llevaban un tiempo sin tener ninguna asociación entre detecciones de *frames* anteriores.

En la sección 5.4 se han incluido cada uno de los valores empíricos determinados de los parámetros que incluye cada uno de los módulos del sistema base.

Por último, en la sección 5.5 se han incluido los resultados obtenidos para cada una de las configuraciones detector-*tracker* que se han utilizado en el proyecto. Como era de esperar, los resultados obtenidos descendían con el aumento de la complejidad de los niveles del *dataset*. Para añadir mejoras al sistema base que se ha desarrollado se ha decidido tomar la decisión de utilizar la combinación del detector de personas Latent SVM y el *tracker* Filtro de Partículas debido a que en el último nivel de complejidad del *dataset* se ha observado un deterioro considerable de los resultados con el Filtro de Kalman.

# 6 Modificaciones del sistema base

---

## 6.1 Introducción

En este capítulo se presentan las mejoras introducidas (sección 6.2) al sistema base implementado para el seguimiento de múltiples personas basado en detección. En cada una de ellas, se realiza una breve descripción de su implementación y se incluye los resultados obtenidos tras la modificación, así como, una comparativa de resultados para medir la mejora introducida con respecto al sistema anterior. Además, se han extraído una serie de conclusiones finales del algoritmo final desarrollado (sección 6.3).

## 6.2 Modificaciones del módulo de matching básico del sistema base

### 6.2.1 Primera modificación del sistema base

#### 6.2.1.1 Actualización del modelo del tracker utilizando únicamente la información del detector de personas

En la sección 4.3 se comentó que el Filtro de Kalman corregía la predicción realizada en cada instante de tiempo utilizando la información de la medida de observación si se disponía de ella y, en la sección 4.4 se explicó que el Filtro de Partículas poseía un paso opcional de actualización del modelo creado en el proceso de inicialización del filtro.

Como se detalló en la sección 5.3.2.2, el *bounding box* de un objetivo con detección asociada a la salida final del sistema venía determinado por la siguiente relación:

$$BB_{final} = (1 - \alpha) * BB_{det} + \alpha * BB_{tr} \quad \text{Ecuación 53}$$

En el sistema base se utilizaban las coordenadas de este *bounding box* para corregir los modelos del *tracker* utilizado en la arquitectura del sistema. Sin embargo, los experimentos realizados muestran que utilizar únicamente la información proporcionada por la nueva detección asociada produce mejores resultados. Esto puede ser debido a que de esta forma se aprovecha la gran precisión que poseen los detectores de personas y se evitan los problemas de que el *tracker* haya podido perder precisión en el seguimiento del objetivo y de que se actualice el modelo del *tracker* con información del modelo de fondo de la escena.

#### 6.2.1.2 Resultados obtenidos para cada nivel del dataset

La Tabla 6-1 muestra los resultados medios obtenidos de las secuencias incluidas en cada uno de los niveles del *dataset* para el detector Latent SVM y el *tracker* Filtro de Partículas tras la primera mejora introducida al sistema base. Los resultados obtenidos para cada una de las secuencias de cada uno de los niveles del *dataset* se han incluido en el Anexo 0 (subapartado B.1).

Métrica	Resultados Nivel 1	Resultados Nivel 2	Resultados Nivel 3	Resultados Nivel 4	Resultados Nivel 5	Resultados Nivel 6
MOTP	0.73574	0.73905	0.71235	0.70220	0.68397	0.693553
MOTA	0.83284	0.68932	0.67425	0.51524	0.28870	0.321717
FP	0.09726	0.09809	0.03458	0.20699	0.25259	0.189227
TP	0.93010	0.78741	0.70883	0.72222	0.54129	0.51094
FN	0.06989	0.21259	0.28980	0.26248	0.45688	0.47391
ID switch	0.00	0.00	0.80	3.00	1.00	9.00

**Tabla 6-1: Resultados medios obtenidos para cada nivel del *dataset* tras la primera mejora introducida al sistema base.**

### 6.2.1.3 Comparativa de resultados

La Tabla 6-2 y la Tabla 6-3 muestran una comparativa de los resultados medios obtenidos en cada uno de los niveles del *dataset* para el detector de personas Latent SVM y el *tracker* Filtro de Partículas, tras introducirse la primera mejora en el sistema base, así como el porcentaje de mejora que se introduce tras la modificación. En negrita se han resaltado los resultados medios que obtienen mejores porcentajes de rendimiento, ya sea antes o después de la mejora introducida.

Como se puede apreciar, los resultados obtenidos tras la primera modificación introducida en el sistema base mejoran tanto para MOTP como para MOTA en cada uno de los niveles del *dataset*, a excepción del primer nivel donde MOTA desciende un 0.3% aproximadamente. Esto se debe a lo comentado anteriormente y es que con la mejora introducida se está aprovechando la alta precisión de los detectores de personas y se está evitando introducir información del modelo de fondo de la escena en el modelo del *tracker* en aquellos casos donde el *tracker* ha podido perder precisión en el proceso de seguimiento del objetivo. Como se puede observar, en la mayoría de niveles la mejora introducida en MOTP es de aproximadamente un 1%, mientras que MOTA incrementa alrededor de un 3% para los niveles del 2 al 4 y de un 8% para los niveles 5 y 6 debido principalmente al incremento del volumen de personas en las secuencias que contienen. Por otro lado, el descenso leve que se produce en MOTA en el primer nivel, se debe a que en un momento determinado en una de las secuencias contenidas en este nivel del *dataset*, el detector de personas genera un falso positivo cercano a uno de los objetivos presentes en la escena, que se asocia con él debido a que supera los umbrales de asociación definidos en el sistema (explicados en el módulo de *matching* en la sección 5.3.2) y para ese objetivo el detector no produce un *bounding box* de salida que le contenga.

Métrica		Nivel 1			
	Resultados etapa anterior	Resultados tras mejora	Mejora incremental tras la modificación	Resultados sist.base	Mejora total sobre el sist.base
MOTP	0.72862	<b>0.73574</b>	0.98%	0.72862	0.98%
MOTA	<b>0.83606</b>	0.83284	-0.39%	0.83606	-0.39%
FP	0.09911	0.09726	1.87%	0.09911	1.87%
TP	0.93517	0.93010	-0.54%	0.93517	-0.54%
FN	0.06483	0.06989	-7.81%	0.06483	-7.81%
ID switch	0.00	0.00	0.00%	0.00	0.00%

Métrica		Nivel 2			
	Resultados etapa anterior	Resultados tras mejora	Mejora incremental tras la modificación	Resultados sist.base	Mejora total sobre el sist.base
MOTP	0.72540	<b>0.73905</b>	1.88%	0.72540	1.88%
MOTA	0.65629	<b>0.68932</b>	5.03%	0.65629	5.03%
FP	0.11495	0.09809	14.67%	0.11495	14.67%
TP	0.77138	0.78741	2.08%	0.77138	2.08%
FN	0.22813	0.21259	6.81%	0.22813	6.81%
ID switch	0.33	0.00	100.00%	0.33	100.00%

Métrica		Nivel 3			
	Resultados etapa anterior	Resultados tras mejora	Mejora incremental tras la modificación	Resultados sist.base	Mejora total sobre el sist.base
MOTP	0.70517	<b>0.71235</b>	1.02%	0.70517	1.02%
MOTA	0.64020	<b>0.67425</b>	5.32%	0.64020	5.32%
FP	0.04334	0.03458	20.21%	0.04334	20.21%
TP	0.68357	0.70883	3.70%	0.68357	3.70%
FN	0.31578	0.28980	8.23%	0.31578	8.23%
ID switch	0.60	0.80	-33.33%	0.60	-33.33%

**Tabla 6-2: Comparativa de resultados medios obtenidos en los 3 primeros niveles del dataset tras la primera mejora introducida con respecto al sistema base. Resultados etapa anterior se corresponde con los resultados obtenidos para el sistema base, Resultados tras mejora hace referencia a los resultados obtenidos tras la mejora introducida, Mejora incremental tras la modificación es el porcentaje de mejora con respecto a la mejora anterior (en este caso, el sistema base), Resultados sistema base son los resultados obtenidos del sistema base original y Mejora total sobre el sistema base es el porcentaje de mejora respecto a los resultados obtenidos del sistema base original, esto es, la acumulación de mejoras.**

Métrica		Nivel 4			
	Resultados etapa anterior	Resultados tras mejora	Mejora incremental tras la modificación	Resultados sist.base	Mejora total sobre el sist.base
MOTP	0.69751	<b>0.70220</b>	0.67%	0.69751	0.67%
MOTA	0.47343	<b>0.51524</b>	8.83%	0.47343	8.83%
FP	0.22339	0.20699	7.34%	0.22339	7.34%
TP	0.69681	0.72222	3.65%	0.69681	3.65%
FN	0.29414	0.26248	10.76%	0.29414	10.76%
ID switch	2.33	3.00	-28.76%	2.33	-28.76%

Métrica		Nivel 5			
	Resultados etapa anterior	Resultados tras mejora	Mejora incremental tras la modificación	Resultados sist.base	Mejora total sobre el sist.base
MOTP	0.66426	<b>0.68397</b>	2.97%	0.66426	2.97%
MOTA	0.20504	<b>0.28870</b>	40.80%	0.20504	40.80%
FP	0.27978	0.25259	9.72%	0.27978	9.72%
TP	0.48481	0.54129	11.65%	0.48481	11.65%
FN	0.51152	0.45688	10.68%	0.51152	10.68%
ID switch	2.00	1.00	50.00%	2.00	50.00%

Métrica		Nivel 6			
	Resultados etapa anterior	Resultados tras mejora	Mejora incremental tras la modificación	Resultados sist.base	Mejora total sobre el sist.base
MOTP	0.68225	<b>0.693553</b>	1.66%	0.68225	1.66%
MOTA	0.23450	<b>0.321717</b>	37.19%	0.23450	37.19%
FP	0.21602	0.189227	12.40%	0.21602	12.40%
TP	0.45052	0.51094	13.41%	0.45052	13.41%
FN	0.53416	0.47391	11.28%	0.53416	11.28%
ID switch	9.33	9.00	3.54%	9.33	3.54%

**Tabla 6-3: Comparativa de resultados medios obtenidos en los 3 últimos niveles del *dataset* tras la primera mejora introducida con respecto al sistema base. Resultados etapa anterior se corresponde con los resultados obtenidos para el sistema base, Resultados tras mejora hace referencia a los resultados obtenidos tras la mejora introducida, Mejora incremental tras la modificación es el porcentaje de mejora con respecto a la mejora anterior (en este caso, el sistema base), Resultados sistema base son los resultados obtenidos del sistema base original y Mejora total sobre el sistema base es el porcentaje de mejora respecto a los resultados obtenidos del sistema base original, esto es, la acumulación de mejoras.**

## 6.2.2 Segunda modificación del sistema base

### 6.2.2.1 Inicialización del tracker al detectarse que pierde al objetivo

En la sección 5.3.2.1 se ha detallado el proceso de inicialización utilizado en el sistema base, según el cual, una nueva detección en el *frame* actual que no se asocia con

detecciones de *frames* anteriores, inicializa una instancia del *tracker* utilizado en el sistema y se asigna al objetivo un ID único y característico.

Uno de los principales problemas que pueden surgir a la hora de utilizar un *tracker* para el seguimiento de personas es la pérdida del objetivo que puede sufrir dicho *tracker*. Para afrontar este problema se ha propuesto inicializar el *tracker* utilizado en el sistema cada vez que no exista el suficiente solapamiento entre la nueva detección asociada y la predicción del *tracker* del objetivo asociado. Se ha definido un umbral  $\beta$  que se ha fijado en un principio a un valor de 0.3. Por tanto, si el porcentaje de solapamiento entre la nueva detección asociada y la predicción del *tracker* del objetivo asociado es menor del 30%, se inicializa el *tracker* haciendo uso del *bounding box* proporcionado por la salida del detector de personas.

### 6.2.2.2 Resultados obtenidos para cada nivel del dataset

La Tabla 6-4 muestra los resultados medios obtenidos de las secuencias incluidas en cada uno de los niveles del *dataset* para el detector Latent SVM y el *tracker* Filtro de Partículas tras la segunda mejora introducida al sistema base. Los resultados obtenidos para cada una de las secuencias de cada uno de los niveles del *dataset* se han incluido en el Anexo 0 (subapartado B.2).

Métrica	Resultados Nivel 1	Resultados Nivel 2	Resultados Nivel 3	Resultados Nivel 4	Resultados Nivel 5	Resultados Nivel 6
MOTP	0.73574	0.72941	0.70213	0.69539	0.68090	0.68990
MOTA	0.83284	0.67351	0.68332	0.53565	0.30170	0.38994
FP	0.09726	0.10624	0.03004	0.19792	0.24196	0.17185
TP	0.93010	0.77975	0.71337	0.73356	0.54366	0.56178
FN	0.06989	0.22025	0.28526	0.24661	0.45359	0.42500
ID switch	0.00	0.00	0.80	3.67	1.50	10.67

**Tabla 6-4: Resultados medios obtenidos para cada nivel del *dataset* tras la segunda mejora introducida al sistema base.**

### 6.2.2.3 Comparativa de resultados

La Tabla 6-5 y la Tabla 6-6 muestran una comparativa de los resultados medios obtenidos en cada uno de los niveles del *dataset* para el detector de personas Latent SVM y el *tracker* Filtro de Partículas, tras introducirse la segunda mejora, así como el porcentaje de mejora que se introduce tras la modificación y el porcentaje de mejora acumulado respecto al sistema base. En negrita se han resaltado los resultados medios que obtienen mejores porcentajes de rendimiento, ya sea antes o después de la mejora introducida.

Como se puede apreciar, los resultados obtenidos tras la segunda modificación introducida empeoran levemente los porcentajes de MOTP. Sin embargo, esta disminución no supera el 1% en ninguno de los niveles del *dataset*. Por otro lado, los resultados de MOTA incrementan para la mayoría de los niveles, a excepción del nivel 1 que ni mejora ni empeora y del nivel 2 que desciende en un 1.5% aproximadamente. El descenso leve que se produce en MOTP puede deberse a que MOTA aumenta en la mayoría de niveles

produciendo un mayor número de *true positives* pudiendo ser la salida del detector de personas algo menos precisa en las nuevas asociaciones que se producen. Por otra parte, el incremento de MOTA se debe a lo comentado en el apartado 6.2.2.1 de esta misma sección y es que con la mejora introducida se reducen las pérdidas de precisión que puede sufrir el *tracker* utilizado a lo largo del seguimiento de un objetivo.

Métrica		Nivel 1			
	Resultados etapa anterior	Resultados tras mejora	Mejora incremental tras la modificación	Resultados sist.base	Mejora total sobre el sist.base
MOTP	0.73574	0.73574	0.00%	0.72862	0.98%
MOTA	0.83284	0.83284	0.00%	0.83606	-0.39%
FP	0.09726	0.09726	0.00%	0.09911	1.87%
TP	0.93010	0.93010	0.00%	0.93517	-0.54%
FN	0.06989	0.06989	0.00%	0.06483	-7.81%
ID switch	0.00	0.00	0.00%	0.00	0.00%

Métrica		Nivel 2			
	Resultados etapa anterior	Resultados tras mejora	Mejora incremental tras la modificación	Resultados sist.base	Mejora total sobre el sist.base
MOTP	<b>0.73905</b>	0.72941	-1.30%	0.72540	0.55%
MOTA	<b>0.68932</b>	0.67351	-2.29%	0.65629	2.62%
FP	0.09809	0.10624	-8.31%	0.11495	7.58%
TP	0.78741	0.77975	-0.97%	0.77138	1.09%
FN	0.21259	0.22025	-3.60%	0.22813	3.45%
ID switch	0.00	0.00	0.00%	0.33	100.00%

Métrica		Nivel 3			
	Resultados etapa anterior	Resultados tras mejora	Mejora incremental tras la modificación	Resultados sist.base	Mejora total sobre el sist.base
MOTP	<b>0.71235</b>	0.70213	-1.43%	0.70517	-0.43%
MOTA	0.67425	<b>0.68332</b>	1.35%	0.64020	6.74%
FP	0.03458	0.03004	13.13%	0.04334	30.69%
TP	0.70883	0.71337	0.64%	0.68357	4.36%
FN	0.28980	0.28526	1.57%	0.31578	9.66%
ID switch	0.80	0.80	0.00%	0.60	-33.33%

**Tabla 6-5: Comparativa de resultados medios obtenidos en los 3 primeros niveles del *dataset* tras la segunda mejora introducida con respecto al sistema base y al sistema base modificado con las mejoras anteriores. Resultados etapa anterior se corresponde con los resultados obtenidos para el sistema base, Resultados tras mejora hace referencia a los resultados obtenidos tras la mejora introducida, Mejora incremental tras la modificación es el porcentaje de mejora con respecto a la mejora anterior (en este caso, el sistema base), Resultados sistema base son los resultados obtenidos del sistema base original y Mejora total sobre el sistema base es el porcentaje de mejora respecto a los resultados obtenidos del sistema base original, esto es, la acumulación de mejoras.**

Métrica		Nivel 4			
	Resultados etapa anterior	Resultados tras mejora	Mejora incremental tras la modificación	Resultados sist.base	Mejora total sobre el sist.base
MOTP	<b>0.70220</b>	0.69539	-0.97%	0.69751	-0.30%
MOTA	0.51524	<b>0.53565</b>	3.96%	0.47343	13.14%
FP	0.20699	0.19792	4.38%	0.22339	11.40%
TP	0.72222	0.73356	1.57%	0.69681	5.27%
FN	0.26248	0.24661	6.05%	0.29414	16.16%
ID switch	3.00	3.67	-22.33%	2.33	-57.51%

Métrica		Nivel 5			
	Resultados etapa anterior	Resultados tras mejora	Mejora incremental tras la modificación	Resultados sist.base	Mejora total sobre el sist.base
MOTP	<b>0.68397</b>	0.68090	-0.45%	0.66426	2.51%
MOTA	0.28870	<b>0.30170</b>	4.50%	0.20504	47.14%
FP	0.25259	0.24196	4.21%	0.27978	13.52%
TP	0.54129	0.54366	0.44%	0.48481	12.14%
FN	0.45688	0.45359	0.72%	0.51152	11.33%
ID switch	1.00	1.50	-50.00%	2.00	25.00%

Métrica		Nivel 6			
	Resultados etapa anterior	Resultados tras mejora	Mejora incremental tras la modificación	Resultados sist.base	Mejora total sobre el sist.base
MOTP	<b>0.69355</b>	0.68990	-0.53%	0.68225	1.21%
MOTA	0.32172	<b>0.38994</b>	21.21%	0.23450	66.29%
FP	0.18923	0.17185	9.18%	0.21602	20.45%
TP	0.51094	0.56178	9.95%	0.45052	24.70%
FN	0.47391	0.42500	10.32%	0.53416	20.44%
ID switch	9.00	10.67	-18.56%	9.33	-14.36%

**Tabla 6-6: Comparativa de resultados medios obtenidos en los 3 últimos niveles del *dataset* tras la segunda mejora introducida con respecto al sistema base y al sistema base modificado con las mejoras anteriores. Resultados etapa anterior se corresponde con los resultados obtenidos para el sistema base, Resultados tras mejora hace referencia a los resultados obtenidos tras la mejora introducida, Mejora incremental tras la modificación es el porcentaje de mejora con respecto a la mejora anterior (en este caso, el sistema base), Resultados sistema base son los resultados obtenidos del sistema base original y Mejora total sobre el sistema base es el porcentaje de mejora respecto a los resultados obtenidos del sistema base original, esto es, la acumulación de mejoras.**

## 6.2.3 Tercera modificación del sistema base

### 6.2.3.1 Inicialización directa de objetivos con detecciones de alta confianza

En la sección 5.3.2.1 se ha comentado el procedimiento que se ha seguido para eliminar en gran medida los falsos negativos que genera un algoritmo de detección de personas y que degradan notablemente los resultados del sistema de seguimiento. Para ello, se había definido un parámetro denominado MIN\_ACTIVE mediante el cual el sistema final no obtenía salida visual del objetivo hasta que al menos se hubiesen realizado el número de asociaciones indicado por este parámetro entre detecciones de *frames* consecutivos.

Para mejorar los resultados del sistema de seguimiento se ha propuesto añadir una mejora al proceso de inicialización del sistema base descrito anteriormente mediante la inicialización directa de objetivos que han sido detectados con una alta confianza. Para ello, se ha definido un umbral  $\mu$  que se ha fijado a 2.5 para el detector de personas utilizado en el sistema base (Latent SVM) tras diversas pruebas experimentales. Por tanto, todas las nuevas detecciones que no se asocien con ningún objetivo existente en el sistema de seguimiento en el *frame* actual y que estén por encima de ese umbral, se inicializan directamente obteniéndose inmediatamente una salida visual y objetiva con la información relevante del objetivo (coordenadas del bounding box, ID característico, etc.) que se almacena en el archivo de resultados descrito en la sección 5.3.2.4.

### 6.2.3.1 Resultados obtenidos para cada nivel del dataset

La Tabla 6-7 muestra los resultados medios obtenidos de las secuencias incluidas en cada uno de los niveles del *dataset* para el detector Latent SVM y el *tracker* Filtro de Partículas tras la tercera mejora introducida al sistema base. Los resultados obtenidos para cada una de las secuencias de cada uno de los niveles del *dataset* se han incluido en el Anexo 0 (subapartado B.1).

Métrica	Resultados Nivel 1	Resultados Nivel 2	Resultados Nivel 3	Resultados Nivel 4	Resultados Nivel 5	Resultados Nivel 6
MOTP	0.73674	0.74356	0.70160	0.69556	0.67840	0.68847
MOTA	0.84443	0.68950	0.69729	0.54806	0.29441	0.40512
FP	0.09726	0.10624	0.03089	0.19792	0.26576	0.18334
TP	0.94168	0.79574	0.72819	0.74598	0.56017	0.58846
FN	0.05832	0.20426	0.27045	0.23363	0.43708	0.39727
ID switch	0.00	0.00	0.80	4.00	1.50	8.67

**Tabla 6-7: Resultados medios obtenidos para cada nivel del *dataset* tras la tercera mejora introducida al sistema base.**

### 6.2.3.2 Comparativa de resultados

La Tabla 6-8 y la Tabla 6-9 muestran una comparativa de los resultados medios obtenidos en cada uno de los niveles del *dataset* para el detector de personas Latent SVM y el *tracker* Filtro de Partículas, tras introducirse la tercera mejora, así como el porcentaje de mejora que se introduce tras la modificación y el porcentaje de mejora acumulado respecto al

sistema base. En **negrita** se han resaltado los resultados medios que obtienen mejores porcentajes de rendimiento, ya sea antes o después de la mejora introducida.

Como se puede apreciar, los resultados obtenidos tras la tercera modificación introducida mejoran o empeoran levemente los resultados de MOTP en función de la precisión del detector de personas en las nuevas asociaciones que se producen al inicializar directamente las detecciones de alta confianza. Por otro lado, se puede observar que los resultados de MOTA aumentan en la mayoría de niveles del *dataset* en un 1.5% aproximadamente, con excepción del nivel 5 que empeora alrededor de un 0.7% aproximadamente y del nivel 1 que mejora levemente en un 0.1%. El descenso que se produce de MOTA en el nivel 5 viene ocasionado porque en una de las secuencias se inicializa un falso positivo que posee una alta confianza produciendo falsos positivos durante los *frames* en los que se permitía utilizar la información del *tracker* sin tener detección asociada del detector de personas y que venía dado por el parámetro MAX\_NO\_DETECTION (explicado en la sección 5.3.2.2).

Métrica		Nivel 1			
	Resultados etapa anterior	Resultados tras mejora	Mejora incremental tras la modificación	Resultados sist.base	Mejora total sobre el sist.base
MOTP	0.73574	<b>0.73674</b>	0.13%	0.72862	1.11%
MOTA	0.83284	<b>0.84443</b>	1.39%	0.83606	1.00%
FP	0.09726	0.09726	0.00%	0.09911	1.87%
TP	0.93010	0.94168	1.25%	0.93517	0.70%
FN	0.06989	0.05832	16.55%	0.06483	10.04%
ID switch	0.00	0.00	0.00%	0.0	0.00%

Métrica		Nivel 2			
	Resultados etapa anterior	Resultados tras mejora	Mejora incremental tras la modificación	Resultados sist.base	Mejora total sobre el sist.base
MOTP	0.72941	<b>0.74356</b>	1.94%	0.72540	2.50%
MOTA	0.67351	<b>0.68950</b>	2.37%	0.65629	5.06%
FP	0.10624	0.10624	0.00%	0.11495	7.58%
TP	0.77975	0.79574	2.05%	0.77138	3.16%
FN	0.22025	0.20426	7.26%	0.22813	10.46%
ID switch	0.00	0.00	0.00%	0.33	100.00%

Métrica		Nivel 3			
	Resultados etapa anterior	Resultados tras mejora	Mejora incremental tras la modificación	Resultados sist.base	Mejora total sobre el sist.base
MOTP	<b>0.70213</b>	0.70160	-0.08%	0.70517	-0.51%
MOTA	0.68332	<b>0.69729</b>	2.04%	0.64020	8.92%
FP	0.03004	0.03089	-2.83%	0.04334	28.73%
TP	0.71337	0.72819	2.08%	0.68357	6.53%
FN	0.28526	0.27045	5.19%	0.31578	14.35%
ID switch	0.80	0.80	0.00%	0.60	-33.33%

**Tabla 6-8: Comparativa de resultados medios obtenidos en los 3 primeros niveles del dataset tras la tercera mejora introducida con respecto al sistema base y al sistema base modificado con las mejoras anteriores. Resultados etapa anterior se corresponde con los resultados obtenidos para el sistema base, Resultados tras mejora hace referencia a los resultados obtenidos tras la mejora introducida, Mejora incremental tras la modificación es el porcentaje de mejora con respecto a la mejora anterior (en este caso, el sistema base), Resultados sistema base son los resultados obtenidos del sistema base original y Mejora total sobre el sistema base es el porcentaje de mejora respecto a los resultados obtenidos del sistema base original, esto es, la acumulación de mejoras.**

Métrica		Nivel 4			
	Resultados etapa anterior	Resultados tras mejora	Mejora incremental tras la modificación	Resultados sist.base	Mejora total sobre el sist.base
MOTP	0.69539	<b>0.69556</b>	0.02%	0.69751	-0.28%
MOTA	0.53565	<b>0.54806</b>	2.32%	0.47343	15.76%
FP	0.19792	0.19792	0.00%	0.22339	11.40%
TP	0.73356	0.74598	1.69%	0.69681	7.06%
FN	0.24661	0.23363	5.26%	0.29414	20.57%
ID switch	3.67	4.00	-8.99%	2.33	-71.67%

Métrica		Nivel 5			
	Resultados etapa anterior	Resultados tras mejora	Mejora incremental tras la modificación	Resultados sist.base	Mejora total sobre el sist.base
MOTP	<b>0.68090</b>	0.67840	-0.37%	0.66426	2.13%
MOTA	<b>0.30170</b>	0.29441	-2.42%	0.20504	43.59%
FP	0.24196	0.26576	-9.84%	0.27978	5.01%
TP	0.54366	0.56017	3.04%	0.48481	15.54%
FN	0.45359	0.43708	3.64%	0.51152	14.55%
ID switch	1.50	1.50	0.00%	2.00	25.00%

Métrica		Nivel 6			
	Resultados etapa anterior	Resultados tras mejora	Mejora incremental tras la modificación	Resultados sist.base	Mejora total sobre el sist.base
MOTP	<b>0.68990</b>	0.68847	-0.21%	0.68225	0.91%
MOTA	0.38994	<b>0.40512</b>	3.89%	0.23450	72.76%
FP	0.17185	0.18334	-6.69%	0.21602	15.13%
TP	0.56178	0.58846	4.75%	0.45052	30.62%
FN	0.42500	0.39727	6.52%	0.53416	25.63%
ID switch	10.67	8.67	18.74%	9.33	7.07%

**Tabla 6-9: Comparativa de resultados medios obtenidos en los 3 últimos niveles del *dataset* tras la tercera mejora introducida con respecto al sistema base y al sistema base modificado con las mejoras anteriores. Resultados etapa anterior se corresponde con los resultados obtenidos para el sistema base, Resultados tras mejora hace referencia a los resultados obtenidos tras la mejora introducida, Mejora incremental tras la modificación es el porcentaje de mejora con respecto a la mejora anterior (en este caso, el sistema base), Resultados sistema base son los resultados obtenidos del sistema base original y Mejora total sobre el sistema base es el porcentaje de mejora respecto a los resultados obtenidos del sistema base original, esto es, la acumulación de mejoras.**

## 6.2.4 Cuarta modificación del sistema base

### 6.2.4.1 Búsqueda del parámetro óptimo $\beta$

En la sección 6.2.2.1 se ha explicado el método que se ha utilizado para evitar en gran medida las pérdidas de precisión del objetivo que se pueden producir en el *tracker* durante el proceso de seguimiento a lo largo de la secuencia de vídeo. Para ello, se inicializaba el *tracker* utilizado en el sistema cada vez que no existía el suficiente solapamiento entre la nueva detección asociada y la predicción del *tracker* asociado. El umbral de solapamiento mínimo venía determinado por el parámetro definido en el sistema  $\beta$ , que en un principio se había fijado a 0.3.

Para mejorar los resultados del sistema de seguimiento implementado se propone en este punto buscar el valor “óptimo” del parámetro  $\beta$  mediante la variación de dicho valor en diversas pruebas experimentales. A continuación, se muestran dos gráficas con los resultados obtenidos para las dos métricas que se han utilizado para medir objetivamente el rendimiento del sistema de seguimiento. Estas gráficas se han obtenido a partir de la comparativa de resultados medios obtenidos para los diferentes valores de  $\beta$  en cada uno de los niveles del *dataset* que se ha incluido en el Anexo 0 (subapartado B.4.1).

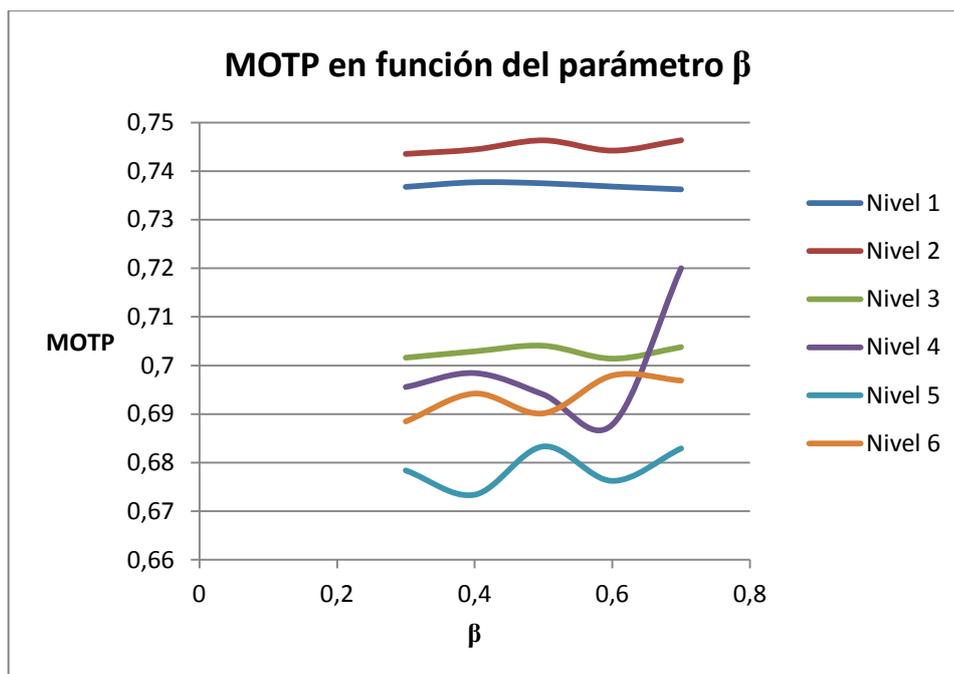
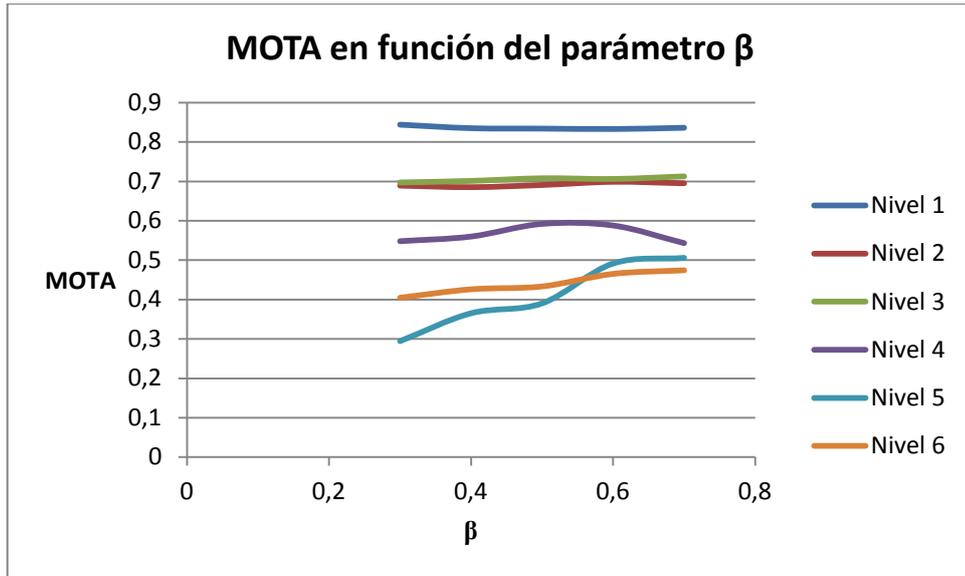


Figura 6-1: MOTP para diferentes valores del parámetro  $\beta$



**Figura 6-2: MOTA para diferentes valores del parámetro  $\beta$**

Observando las gráficas anteriores se ha decidido fijar el parámetro  $\beta$  a un valor “óptimo” de 0.6. Esta decisión se basa en que MOTP obtiene unos resultados parecidos para cada valor de  $\beta$  en cada nivel y MOTA mejora considerablemente para los niveles más complicados del *dataset* creado.

#### 6.2.4.1 Resultados obtenidos para cada nivel del dataset

La Tabla 6-10 muestra los resultados medios obtenidos de las secuencias incluidas en cada uno de los niveles del *dataset* para el detector Latent SVM y el *tracker* Filtro de Partículas tras la cuarta mejora introducida al sistema base para un valor del parámetro  $\beta$  de 0.6. Los resultados obtenidos para cada una de las secuencias de cada uno de los niveles del *dataset* para cada parámetro  $\beta$  probado se han incluido en el Anexo 0 (subapartado 0).

Métrica	Resultados Nivel 1	Resultados Nivel 2	Resultados Nivel 3	Resultados Nivel 4	Resultados Nivel 5	Resultados Nivel 6
MOTP	0.73681	0.74423	0.70140	0.68781	0.67626	0.69793
MOTA	0.83316	0.69930	0.70642	0.58815	0.49170	0.46572
FP	0.10289	0.10133	0.02497	0.14166	0.16774	0.14506
TP	0.93605	0.80064	0.73139	0.72981	0.65944	0.61078
FN	0.06400	0.19936	0.26724	0.26226	0.33782	0.37353
ID switch	0.00	0.00	0.80	1.67	1.50	9.33

**Tabla 6-10: Resultados medios obtenidos para cada nivel del *dataset* tras la cuarta mejora introducida al sistema base.**

#### 6.2.4.2 Comparativa de resultados

La Tabla 6-11 y la Tabla 6-12 muestran una comparativa de los resultados medios obtenidos en cada uno de los niveles del *dataset* para el detector de personas Latent SVM y el *tracker* Filtro de Partículas, tras introducirse la cuarta mejora, así como el porcentaje de mejora que se introduce tras la modificación y el porcentaje de mejora acumulado respecto al sistema base. En negrita se han resaltado los resultados medios que obtienen mejores porcentajes de rendimiento, ya sea antes o después de la mejora introducida.

Como se puede apreciar, los resultados obtenidos tras la cuarta modificación introducida mejoran en prácticamente todos los niveles del *dataset* los resultados de MOTA, a excepción del nivel 1 donde disminuye un 1%. En los niveles más complicados (cinco y seis) se observa una gran mejora producida ya que el volumen de personas en la escena incrementa considerablemente y por tanto, también las pérdidas de precisión del *tracker* en el seguimiento del objetivo. Por ello, los resultados mejoran considerablemente respecto a la mejora que se introdujo en el punto 6.2.2 donde el parámetro  $\beta$  se había fijado en un principio a un valor de 0.3 ya que incrementando este parámetro se reducen un gran número de falsos positivos al aprovechar la precisión del detector de personas para guiar el *tracker* durante el proceso de seguimiento de un objetivo. Por otro lado, los resultados obtenidos tras la cuarta modificación introducida mejoran o empeoran levemente los resultados de MOTP en función de la precisión del detector de personas en las nuevas asociaciones que se producen con la mejora introducida.

Métrica		Nivel 1			
	Resultados etapa anterior	Resultados tras mejora	Mejora incremental tras la modificación	Resultados sist.base	Mejora total sobre el sist.base
MOTP	0.73674	<b>0.73681</b>	0.01%	0.72862	1.12%
MOTA	<b>0.84443</b>	0.83316	-1.33%	0.83606	-0.35%
FP	0.09726	0.10289	-5.79%	0.09911	-3.81%
TP	0.94168	0.93605	-0.60%	0.93517	0.09%
FN	0.05832	0.06400	-9.74%	0.06483	1.28%
ID switch	0.00	0.00	0.00%	0.00	0.00%

Métrica		Nivel 2			
	Resultados etapa anterior	Resultados tras mejora	Mejora incremental tras la modificación	Resultados sist.base	Mejora total sobre el sist.base
MOTP	0.74356	<b>0.74423</b>	0.09%	0.72540	2.60%
MOTA	0.68950	<b>0.69930</b>	1.42%	0.65629	6.55%
FP	0.10624	0.10133	4.62%	0.11495	11.85%
TP	0.79574	0.80064	0.62%	0.77138	3.79%
FN	0.20426	0.19936	2.40%	0.22813	12.61%
ID switch	0.00	0.00	0.00%	0.33	100.00%

Métrica		Nivel 3			
	Resultados etapa anterior	Resultados tras mejora	Mejora incremental tras la modificación	Resultados sist.base	Mejora total sobre el sist.base
MOTP	<b>0.70160</b>	0.70140	-0.03%	0.70517	-0.53%
MOTA	0.69729	<b>0.70642</b>	1.31%	0.64020	10.34%
FP	0.03089	0.02497	19.16%	0.04334	42.39%
TP	0.72819	0.73139	0.44%	0.68357	7.00%
FN	0.27045	0.26724	1.19%	0.31578	15.37%
ID switch	0.80	0.80	0.00%	0.60	-33.33%

**Tabla 6-11: Comparativa de resultados medios obtenidos en los 3 primeros niveles del dataset tras la cuarta mejora introducida con respecto al sistema base y al sistema base modificado con las mejoras anteriores. Resultados etapa anterior se corresponde con los resultados obtenidos para el sistema base, Resultados tras mejora hace referencia a los resultados obtenidos tras la mejora introducida, Mejora incremental tras la modificación es el porcentaje de mejora con respecto a la mejora anterior (en este caso, el sistema base), Resultados sistema base son los resultados obtenidos del sistema base original y Mejora total sobre el sistema base es el porcentaje de mejora respecto a los resultados obtenidos del sistema base original, esto es, la acumulación de mejoras.**

Métrica		Nivel 4			
	Resultados etapa anterior	Resultados tras mejora	Mejora incremental tras la modificación	Resultados sist.base	Mejora total sobre el sist.base
MOTP	<b>0.69556</b>	0.68781	-1.11%	0.69751	-1.39%
MOTA	0.54806	<b>0.58815</b>	7.31%	0.47343	24.23%
FP	0.19792	0.14166	28.43%	0.22339	36.59%
TP	0.74598	0.72981	-2.17%	0.69681	4.74%
FN	0.23363	0.26226	-12.25%	0.29414	10.84%
ID switch	4.00	1.67	58.25%	2.33	28.33%

Métrica		Nivel 5			
	Resultados etapa anterior	Resultados tras mejora	Mejora incremental tras la modificación	Resultados sist.base	Mejora total sobre el sist.base
MOTP	<b>0.67840</b>	0.67626	-0.31%	0.66426	1.81%
MOTA	0.29441	<b>0.49170</b>	67.01%	0.20504	139.81%
FP	0.26576	0.16774	36.88%	0.27978	40.04%
TP	0.56017	0.65944	17.72%	0.48481	36.02%
FN	0.43708	0.33782	22.71%	0.51152	33.96%
ID switch	1.50	1.50	0.00%	2.00	25.00%

Métrica		Nivel 6			
	Resultados etapa anterior	Resultados tras mejora	Mejora incremental tras la modificación	Resultados sist.base	Mejora total sobre el sist.base
MOTP	0.68847	<b>0.69793</b>	1.37%	0.68225	2.30%
MOTA	0.40512	<b>0.46572</b>	14.96%	0.23450	98.60%
FP	0.18334	0.14506	20.88%	0.21602	32.85%
TP	0.58846	0.61078	3.79%	0.45052	35.57%
FN	0.39727	0.37353	5.98%	0.53416	30.07%
ID switch	8.67	9.33	-7.07%	9.33	0.00%

**Tabla 6-12: Comparativa de resultados medios obtenidos en los 3 últimos niveles del dataset tras la cuarta mejora introducida con respecto al sistema base y al sistema base modificado con las mejoras anteriores. Resultados etapa anterior se corresponde con los resultados obtenidos para el sistema base, Resultados tras mejora hace referencia a los resultados obtenidos tras la mejora introducida, Mejora incremental tras la modificación es el porcentaje de mejora con respecto a la mejora anterior (en este caso, el sistema base), Resultados sistema base son los resultados obtenidos del sistema base original y Mejora total sobre el sistema base es el porcentaje de mejora respecto a los resultados obtenidos del sistema base original, esto es, la acumulación de mejoras.**

## 6.2.5 Quinta modificación del sistema base

### 6.2.5.1 Muerte directa de objetivos en las zonas de salida de la secuencia

En la sección 5.3.2.3 se ha detallado el proceso que se ha seguido para la terminación o muerte de objetivos en la escena. Para ello, se había definido un parámetro `MAX_INACTIVE` mediante el cual se eliminaba un objetivo inactivo de la escena cuando tras el número de *frames* indicados por este parámetro el objetivo no había realizado ninguna asociación con detecciones de *frames* posteriores. Un objetivo se consideraba que se encontraba inactivo en el sistema si como se explicaba en la sección 5.3.2.1 no había superado el número de asociaciones indicado por el parámetro `MIN_ACTIVE` entre detecciones de *frames* consecutivos, y por tanto, aún no había sido inicializado, o si un objetivo activo en el sistema superaba el número de predicciones permitidas sin detección asociada del *tracker* definido por el parámetro `MAX_NO_DETECTION`.

Sin embargo, mediante este proceso de terminación de objetivos se pueden producir falsos positivos en las zonas de la secuencia donde los objetivos salen de la escena debido a que como se ha explicado anteriormente se permite utilizar la información de predicción del *tracker* durante el número de *frames* definido por el parámetro `MAX_NO_DETECTION`. Para afrontar este problema se ha propuesto terminar directamente aquellos objetivos que se encuentren saliendo de la zona de visión de la cámara. Para ello, se han definido zonas de salida en las secuencias del *dataset* propuesto que contienen objetivos que salen de la zona de visión de la cámara. Estas zonas se describen mediante pequeñas áreas rectangulares que siguen el siguiente formato:

$$x \ y \ w \ h$$

donde  $x$  y  $y$  son las coordenadas del vértice izquierdo de la parte superior del área rectangular y  $w$  y  $h$  indican el ancho y la altura del mismo área rectangular, respectivamente.

Para comprobar que un objetivo se encuentra en una de las zonas de salida definidas para una secuencia, se comprueban sus cuatro vértices de su *bounding box*. Si alguno de sus vértices se encuentran situados dentro del área de dicha zona se mata directamente dicho objetivo de tal forma que así, se evita la posible generación de falsos positivos debidos a la utilización de la información del *tracker* sin tener detección asociada.

En la Tabla 6-13 se pueden observar las zonas de salida definidas en las secuencias del *dataset* propuesto que contienen objetivos que salen de la zona de visión de la cámara. Como se puede apreciar sólo se han añadido aquellas secuencias en las que se ha añadido alguna zona de salida.

	Nivel 2	Nivel 4			Nivel 5			Nivel 6			
	Secuencia 3	Secuencia 1	Secuencia 2	Secuencia 3	Secuencia 1	Secuencia 2	Secuencia 1	Secuencia 2	Secuencia 3		
	Zona 1	Zona 1	Zona 2	Zona 1	Zona 1	Zona 2	Zona 1	Zona 1	Zona 1		
x	574	24	0	605	0	753	753	0	753	0	753
y	60	77	280	80	22	196	196	124	196	124	196
w	32	35	20	33	7	15	15	15	15	15	15
h	103	131	295	142	246	369	369	211	369	211	369

**Tabla 6-13: Definición de las zonas de salida en cada una de las secuencias del nivel**

### 6.2.5.2 Resultados obtenidos para cada nivel del dataset

La Tabla 6-14 muestra los resultados medios obtenidos de las secuencias incluidas en cada uno de los niveles del *dataset* para el detector Latent SVM y el *tracker* Filtro de Partículas tras la quinta mejora introducida al sistema base. Los resultados obtenidos para cada una de las secuencias de cada uno de los niveles del *dataset* se han incluido en el Anexo 0 (subapartado B.5).

Métrica	Resultados Nivel 1	Resultados Nivel 2	Resultados Nivel 3	Resultados Nivel 4	Resultados Nivel 5	Resultados Nivel 6
MOTP	0.73681	0.74439	0.70139	0.68781	0.67668	0.69793
MOTA	0.83316	0.70039	0.70642	0.60252	0.51676	0.47427
FP	0.10289	0.09809	0.02497	0.12730	0.13840	0.13651
TP	0.93605	0.79848	0.73139	0.72981	0.65516	0.61078
FN	0.06400	0.20152	0.26724	0.26226	0.34209	0.37353
ID switch	0.00	0.00	0.80	1.67	1.50	9.33

**Tabla 6-14: Resultados medios obtenidos para cada nivel del *dataset* tras la quinta mejora introducida al sistema base.**

### 6.2.5.3 Comparativa de resultados

La Tabla 6-15 y la Tabla 6-16 muestran una comparativa de los resultados medios obtenidos en cada uno de los niveles del *dataset* para el detector de personas Latent SVM y el *tracker* Filtro de Partículas, tras introducirse la quinta mejora, así como el porcentaje de mejora que se introduce tras la modificación y el porcentaje de mejora acumulado respecto al sistema base. En negrita se han resaltado los resultados medios que obtienen mejores porcentajes de rendimiento, ya sea antes o después de la mejora introducida.

Como se puede apreciar, los resultados en MOTP obtenidos tras la quinta modificación introducida prácticamente no se ven alterados con respecto a la modificación anterior. Esto era de esperar ya que con esta mejora lo que se reducen son los falsos positivos que se pueden producir en las zonas de salida de los objetivos en cada una de las secuencias del *dataset*, pero las asociaciones que se producían en la modificación anterior siguen siendo las mismas o prácticamente las mismas dependiendo de si un objetivo se elimina un poco antes de que desaparezca del *ground-truth*. Por otro lado, los resultados de MOTA aumentan en todos los niveles del *dataset* en los que se ha definido alguna zona de salida

en las secuencias. Esto se debe a lo explicado anteriormente y es que de esta forma se reducen los falsos positivos que se pueden producir al utilizar la información de predicción del *tracker* durante el número de *frames* definido por el parámetro *MAX\_NO\_DETECTION*.

Métrica		Nivel 1			
	Resultados etapa anterior	Resultados tras mejora	Mejora incremental tras la modificación	Resultados sist.base	Mejora total sobre el sist.base
MOTP	0.73681	0.73681	0.00%	0.72862	1.12%
MOTA	0.83316	0.83316	0.00%	0.83606	-0.35%
FP	0.10289	0.10289	0.00%	0.09911	-3.81%
TP	0.93605	0.93605	0.00%	0.93517	0.09%
FN	0.06400	0.06400	0.00%	0.06483	1.28%
ID switch	0.00	0.00	0.00%	0.00	0.00%

Métrica		Nivel 2			
	Resultados etapa anterior	Resultados tras mejora	Mejora incremental tras la modificación	Resultados sist.base	Mejora total sobre el sist.base
MOTP	0.74423	<b>0.74439</b>	0.02%	0.72540	2.62%
MOTA	0.69930	<b>0.70039</b>	0.16%	0.65629	6.72%
FP	0.10133	0.09809	3.20%	0.11495	14.67%
TP	0.80064	0.79848	-0.27%	0.77138	3.51%
FN	0.19936	0.20152	-1.08%	0.22813	11.66%
ID switch	0.00	0.00	0.00%	0.33	100.00%

Métrica		Nivel 3			
	Resultados etapa anterior	Resultados tras mejora	Mejora incremental tras la modificación	Resultados sist.base	Mejora total sobre el sist.base
MOTP	0.70140	0.70140	0.00%	0.70517	-0.53%
MOTA	0.70642	0.70642	0.00%	0.64020	10.34%
FP	0.02497	0.02497	0.00%	0.04334	42.39%
TP	0.73139	0.73139	0.00%	0.68357	7.00%
FN	0.26724	0.26724	0.00%	0.31578	15.37%
ID switch	0.80	0.80	0.00%	0.60	-33.33%

**Tabla 6-15: Comparativa de resultados medios obtenidos en los 3 primeros niveles del *dataset* tras la quinta mejora introducida con respecto al sistema base y al sistema base modificado con las mejoras anteriores. Resultados etapa anterior se corresponde con los resultados obtenidos para el sistema base, Resultados tras mejora hace referencia a los resultados obtenidos tras la mejora introducida, Mejora incremental tras la modificación es el porcentaje de mejora con respecto a la mejora anterior (en este caso, el sistema base), Resultados sistema base son los resultados obtenidos del sistema base original y Mejora total sobre el sistema base es el porcentaje de mejora respecto a los resultados obtenidos del sistema base original, esto es, la acumulación de mejoras.**

Métrica		Nivel 4			
	Resultados etapa anterior	Resultados tras mejora	Mejora incremental tras la modificación	Resultados sist.base	Mejora total sobre el sist.base
MOTP	0.68781	0.68781	0.00%	0.69751	-1.39%
MOTA	0.58815	<b>0.60252</b>	2.44%	0.47343	27.27%
FP	0.14166	0.12730	10.14%	0.22339	43.01%
TP	0.72981	0.72981	0.00%	0.69681	4.74%
FN	0.26226	0.26226	0.00%	0.29414	10.84%
ID switch	1.67	1.67	0.00%	2.33	28.33%

Métrica		Nivel 5			
	Resultados etapa anterior	Resultados tras mejora	Mejora incremental tras la modificación	Resultados sist.base	Mejora total sobre el sist.base
MOTP	0.67626	<b>0.67668</b>	0.06%	0.66426	1.87%
MOTA	0.49170	<b>0.51676</b>	5.10%	0.20504	152.03%
FP	0.16774	0.13840	17.49%	0.27978	50.53%
TP	0.65944	0.65516	-0.65%	0.48481	35.14%
FN	0.33782	0.34209	-1.26%	0.51152	33.12%
ID switch	1.50	1.50	0.00%	2.00	25.00%

Métrica		Nivel 6			
	Resultados etapa anterior	Resultados tras mejora	Mejora incremental tras la modificación	Resultados sist.base	Mejora total sobre el sist.base
MOTP	0.69793	0.69793	0.00%	0.68225	2.30%
MOTA	0.46572	<b>0.47427</b>	1.84%	0.23450	102.25%
FP	0.14506	0.13651	5.89%	0.21602	36.81%
TP	0.61078	0.61078	0.00%	0.45052	35.57%
FN	0.37353	0.37353	0.00%	0.53416	30.07%
ID switch	9.33	9.33	0.00%	9.33	0.00%

**Tabla 6-16: Comparativa de resultados medios obtenidos en los 3 últimos niveles del dataset tras la quinta mejora introducida con respecto al sistema base y al sistema base modificado con las mejoras anteriores. Resultados etapa anterior se corresponde con los resultados obtenidos para el sistema base, Resultados tras mejora hace referencia a los resultados obtenidos tras la mejora introducida, Mejora incremental tras la modificación es el porcentaje de mejora con respecto a la mejora anterior (en este caso, el sistema base), Resultados sistema base son los resultados obtenidos del sistema base original y Mejora total sobre el sistema base es el porcentaje de mejora respecto a los resultados obtenidos del sistema base original, esto es, la acumulación de mejoras.**

## 6.2.6 Sexta modificación del sistema base

### 6.2.6.1 Búsqueda del parámetro óptimo $\alpha$

En la sección 5.3.2.2 se ha explicado el método utilizado para combinar la salida del detector de personas y del *tracker* utilizados en el sistema de seguimiento múltiple de personas en el proceso de asociación de identidades llevado a cabo en el módulo de *matching*. Para ello, cuando se producía una asociación entre una nueva detección producida por el detector de personas en el *frame* actual con una predicción del *tracker* de uno de los objetivos presentes en el sistema, el *bounding box* de salida del objetivo venía dado por la siguiente relación:

$$\text{Posicion final} = (1 - \alpha) * \text{Posicion}_{det} + \alpha * \text{Posicion}_{tr} \quad \text{Ecuación 54}$$

Para mejorar los resultados del sistema de seguimiento implementado se propone en este punto buscar el valor “óptimo” del parámetro  $\alpha$  que en un principio se había fijado a un valor de 0.5 mediante la variación de dicho valor en diversas pruebas experimentales. A continuación, se muestran dos gráficas con los resultados obtenidos para las dos métricas que se han utilizado para medir objetivamente el rendimiento del sistema de seguimiento. Estas gráficas se han obtenido a partir de la comparativa de resultados medios obtenidos para los diferentes valores de  $\alpha$  en cada uno de los niveles del *dataset* que se ha incluido en el Anexo 0 (subapartado B.6.1).

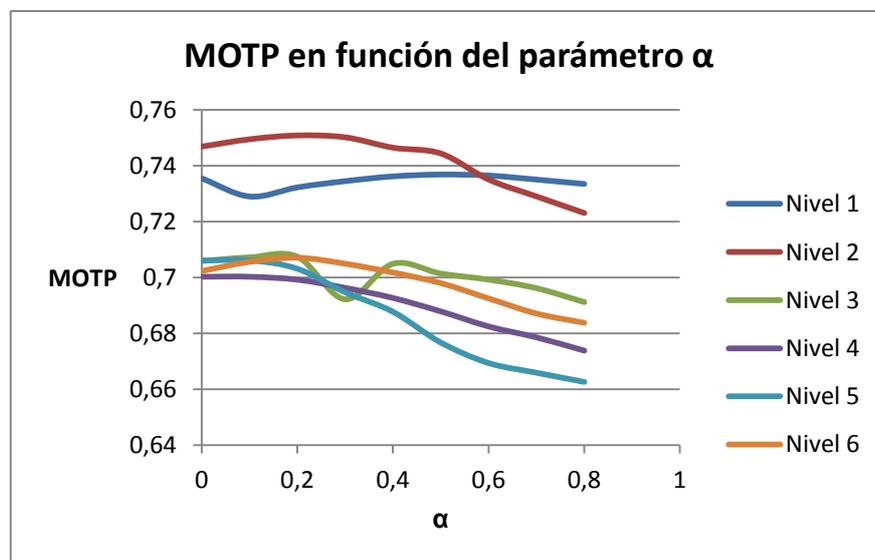
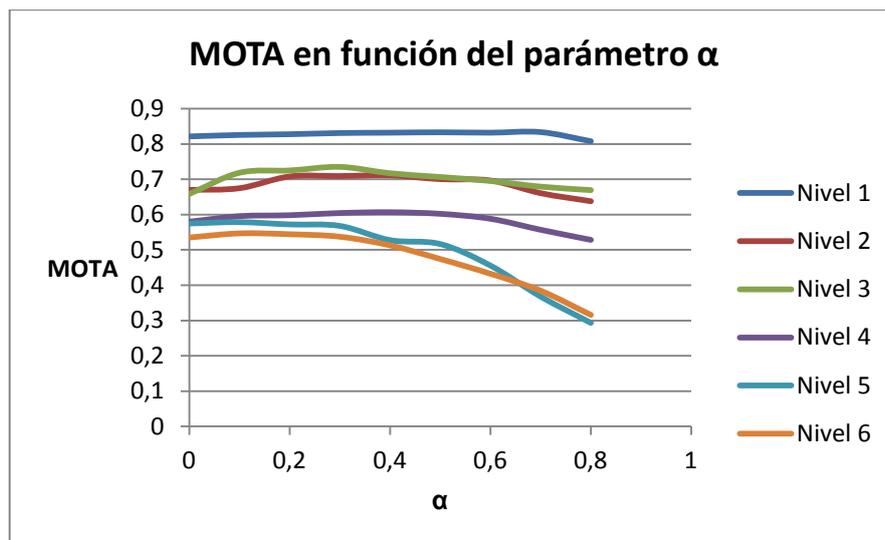


Figura 6-3: MOTP para diferentes valores del parámetro  $\alpha$



**Figura 6-4: MOTA para diferentes valores del parámetro  $\alpha$**

Observando las gráficas anteriores se puede apreciar que tanto MOTP como MOTA mejoran para valores de  $\alpha$  más pequeños. Esto se debe a que los detectores de personas poseen una gran precisión en su proceso de detección. Se ha decidido fijar el parámetro  $\alpha$  a un valor “óptimo” de 0.2 ya que si observamos los resultados que se obtienen para MOTA los resultados descienden levemente para valores de  $\alpha$  más pequeños, mientras que MOTP obtiene resultados muy parecidos en el rango de  $\alpha$  de 0,0 a 0,3.

### 6.2.6.2 Resultados obtenidos para cada nivel del dataset

La Tabla 6-17 muestra los resultados medios obtenidos de las secuencias incluidas en cada uno de los niveles del *dataset* para el detector Latent SVM y el *tracker* Filtro de Partículas tras la sexta mejora introducida al sistema base para un valor del parámetro  $\alpha$  de 0.6. Los resultados obtenidos para cada una de las secuencias de cada uno de los niveles del *dataset* para cada parámetro  $\alpha$  probado se han incluido en el Anexo 0 (subapartado B.1).

Métrica	Resultados Nivel 1	Resultados Nivel 2	Resultados Nivel 3	Resultados Nivel 4	Resultados Nivel 5	Resultados Nivel 6
MOTP	0.73223	0.75085	0.70743	0.69915	0.70312	0.70703
MOTA	0.82776	0.70802	0.72495	0.59860	0.57241	0.54441
FP	0.10559	0.09427	0.01842	0.12926	0.10904	0.10319
TP	0.93335	0.80230	0.74338	0.72785	0.68177	0.64760
FN	0.06665	0.19770	0.25525	0.26422	0.31549	0.33741
ID switch	0.00	0.00	0.80	1.67	1.50	9.00

**Tabla 6-17: Resultados medios obtenidos para cada nivel del *dataset* tras la sexta mejora introducida al sistema base.**

### 6.2.6.3 Comparativa de resultados

La Tabla 6-18 y la Tabla 6-19 muestran una comparativa de los resultados medios obtenidos en cada uno de los niveles del *dataset* para el detector de personas Latent SVM y el *tracker* Filtro de Partículas tras introducirse la sexta mejora, así como el porcentaje de mejora que se introduce tras la modificación y el porcentaje de mejora acumulado respecto al sistema base. En negrita se han resaltado los resultados medios que obtienen mejores porcentajes de rendimiento, ya sea antes o después de la mejora introducida.

Como se puede apreciar, los resultados obtenidos tras la sexta modificación introducida mejoran tanto para MOTP como para MOTA en cada uno de los niveles del *dataset*, a excepción del primer nivel donde MOTA desciende un 0.6% aproximadamente y MOTP alrededor de un 0.4% y del quinto nivel donde MOTA desciende un 0.4%. Sin embargo, este descenso no es comparable al aumento de rendimiento que se consigue en MOTA con la fijación del parámetro  $\alpha$  a 0.2 en los niveles 5 y 6 donde los resultados aumentan un 6% y un 7% aproximadamente. Como se ha explicado anteriormente, que MOTA y MOTP funcionen mejor para valores de  $\alpha$  más pequeños se debe a la gran precisión de los detectores de personas en su proceso de detección. De esta forma al reducir el valor de  $\alpha$  se disminuye el número de falsos positivos y se aumenta el número de asociaciones o de *true positives*.

Métrica		Nivel 1			
	Resultados etapa anterior	Resultados tras mejora	Mejora incremental tras la modificación	Resultados sist.base	Mejora total sobre el sist.base
MOTP	<b>0.73681</b>	0.73223	-0.62%	0.72862	0.50%
MOTA	<b>0.83316</b>	0.82776	-0.64%	0.83606	-0.99%
FP	0.10289	0.10559	-2.62%	0.09911	-6.54%
TP	0.93605	0.93335	-0.29%	0.93517	-0.19%
FN	0.06400	0.06665	-4.14%	0.06483	2.81%
ID switch	0.00	0.00	0.00%	0.00	0.00%

Métrica		Nivel 2			
	Resultados etapa anterior	Resultados tras mejora	Mejora incremental tras la modificación	Resultados sist.base	Mejora total sobre el sist.base
MOTP	0.74439	<b>0.75085</b>	0.87%	0.72540	3.51%
MOTA	0.70039	<b>0.70802</b>	1.09%	0.65629	7.88%
FP	0.09809	0.09427	3.89%	0.11495	17.99%
TP	0.79848	0.80230	0.48%	0.77138	4.01%
FN	0.20152	0.19770	1.90%	0.22813	13.34%
ID switch	0.00	0.00	0.00%	0.33	100.00%

Métrica		Nivel 3			
	Resultados etapa anterior	Resultados tras mejora	Mejora incremental tras la modificación	Resultados sist.base	Mejora total sobre el sist.base
MOTP	0.70139	<b>0.70743</b>	0.86%	0.70517	0.32%
MOTA	0.70642	<b>0.72495</b>	2.62%	0.64020	13.24%
FP	0.02497	0.01842	26.23%	0.04334	57.50%
TP	0.73139	0.74338	1.64%	0.68357	8.75%
FN	0.26724	0.25525	4.49%	0.31578	19.17%
ID switch	0.80	0.80	0.00%	0.60	-33.33%

**Tabla 6-18: Comparativa de resultados medios obtenidos en los 3 primeros niveles del *dataset* tras la sexta mejora introducida con respecto al sistema base y al sistema base modificado con las mejoras anteriores. Resultados etapa anterior se corresponde con los resultados obtenidos para el sistema base, Resultados tras mejora hace referencia a los resultados obtenidos tras la mejora introducida, Mejora incremental tras la modificación es el porcentaje de mejora con respecto a la mejora anterior (en este caso, el sistema base), Resultados sistema base son los resultados obtenidos del sistema base original y Mejora total sobre el sistema base es el porcentaje de mejora respecto a los resultados obtenidos del sistema base original, esto es, la acumulación de mejoras.**

Métrica		Nivel 4			
	Resultados etapa anterior	Resultados tras mejora	Mejora incremental tras la modificación	Resultados sist.base	Mejora total sobre el sist.base
MOTP	0.68781	<b>0.69915</b>	1.65%	0.69751	0.24%
MOTA	<b>0.60252</b>	0.59860	-0.65%	0.47343	26.44%
FP	0.12730	0.12926	-1.54%	0.22339	42.14%
TP	0.72981	0.72785	-0.27%	0.69681	4.45%
FN	0.26226	0.26422	-0.75%	0.29414	10.17%
ID switch	1.67	1.67	0.00%	2.33	28.33%

Métrica		Nivel 5			
	Resultados etapa anterior	Resultados tras mejora	Mejora incremental tras la modificación	Resultados sist.base	Mejora total sobre el sist.base
MOTP	0.67668	<b>0.70312</b>	3.91%	0.66426	5.85%
MOTA	0.51676	<b>0.57241</b>	10.77%	0.20504	179.17%
FP	0.13840	0.10904	21.21%	0.27978	61.03%
TP	0.65516	0.68177	4.06%	0.48481	40.63%
FN	0.34209	0.31549	7.78%	0.51152	38.32%
ID switch	1.50	1.50	0.00%	2.00	25.00%

Métrica		Nivel 6			
	Resultados etapa anterior	Resultados tras mejora	Mejora incremental tras la modificación	Resultados sist.base	Mejora total sobre el sist.base
MOTP	0.69793	<b>0.70703</b>	1.30%	0.68225	3.63%
MOTA	0.47427	<b>0.54441</b>	14.79%	0.23450	132.16%
FP	0.13651	0.10319	24.41%	0.21602	52.23%
TP	0.61078	0.64760	6.03%	0.45052	43.75%
FN	0.37353	0.33741	9.67%	0.53416	36.83%
ID switch	9.33	9.00	4.00%	9.33	3.54%

**Tabla 6-19: Comparativa de resultados medios obtenidos en los 3 últimos niveles del dataset tras la sexta mejora introducida con respecto al sistema base y al sistema base modificado con las mejoras anteriores. Resultados etapa anterior se corresponde con los resultados obtenidos para el sistema base, Resultados tras mejora hace referencia a los resultados obtenidos tras la mejora introducida, Mejora incremental tras la modificación es el porcentaje de mejora con respecto a la mejora anterior (en este caso, el sistema base), Resultados sistema base son los resultados obtenidos del sistema base original y Mejora total sobre el sistema base es el porcentaje de mejora respecto a los resultados obtenidos del sistema base original, esto es, la acumulación de mejoras.**

## 6.3 Conclusiones

Durante este capítulo se han presentado las mejoras introducidas en el sistema base que se ha desarrollado para el seguimiento de múltiples personas en vídeo basado en detección (sección 6.2). Para ello, se ha realizado una breve descripción de cada una de ellas, se han incluido los resultados obtenidos tras cada modificación y se ha realizado una comparativa en cada una de las mejoras introducidas para medir la mejora introducida con respecto al sistema anterior y al sistema base.

La primera mejora introducida (sección 6.2.1) ha sido utilizar únicamente la información proporcionada por la nueva detección asociada para actualizar el modelo del *tracker* utilizado en la arquitectura. Mediante esta mejora se ha aprovechado la gran precisión de los detectores de personas y se ha evitado introducir información del modelo del fondo de la escena en el proceso de actualización del *tracker*.

La segunda mejora introducida (sección 6.2.2) ha sido inicializar el *tracker* utilizado en el sistema cada vez que no existiese el suficiente solapamiento entre la nueva detección asociada y la predicción del *tracker* del objetivo asociado. De esta forma, se solucionaban los problemas de la pérdida del objetivo que puede sufrir dicho *tracker* durante el proceso de seguimiento. Para ello, se ha definido un umbral  $\beta$  que mide el solapamiento entre la detección asociada y la predicción del *tracker*. Además, este parámetro ha sido optimizado en la cuarta mejora introducida (sección 6.2.4) para mejorar los resultados experimentales.

La tercera mejora que se ha propuesto (sección 6.2.3) ha sido la inicialización directa de objetivos que han sido detectados con una alta confianza. Para ello, se ha fijado un umbral mediante el cual todas las nuevas detecciones que no se asocien con ningún objetivo existente en el sistema de seguimiento en el *frame* actual y que estén por encima de ese umbral, se inicializan directamente obteniéndose inmediatamente una salida visual y objetiva con la información relevante del objetivo.

La quinta mejora que se ha añadido (sección 6.2.5) eliminaba falsos positivos que se producían en las zonas de la secuencia donde los objetivos salen de la escena. Estos falsos positivos se producían por el uso de la información de predicción del *tracker* durante un número determinado de *frames*.

La última mejora introducida (sección 6.2.6) ha sido la optimización del parámetro  $\alpha$  definido en el módulo de Matching del sistema base (sección 5.3.2.2) cuyo valor se ha visto reducido del valor inicial debido a la alta precisión del detector utilizado en el sistema base.

Si se observan los resultados finales obtenidos se puede apreciar que en cada uno de los niveles del *dataset* los resultados medios de MOTP mejoran poco con respecto al sistema base implementado inicialmente. Esto se debe a que MOTP mide la precisión de la localización del objetivo durante el proceso del seguimiento y depende únicamente de la precisión del detector utilizado en el sistema. En cambio, MOTA que evalúa el rendimiento a nivel de secuencia del proceso del seguimiento teniendo en cuenta la información de falsos positivos, falsos negativos y cambios de identificador aumenta considerablemente con respecto a los resultados obtenidos para el sistema base (en mayor medida en los niveles más complejos del *dataset*) a excepción del primer nivel del *dataset* donde desciende levemente en un uno por ciento aproximadamente. Este descenso se debe

a que en un momento determinado en una de las secuencias contenidas en este nivel del *dataset*, el detector de personas genera un falso positivo cercano a uno de los objetivos presentes en la escena, que se asocia con él debido a que supera los umbrales de asociación definidos en el sistema. Además, se ha buscado optimizar los parámetros  $\alpha$  y  $\beta$  que para los nuevos valores fijados los resultados obtenidos descendían levemente para este nivel, pero que, en cambio, aumentaban notablemente los resultados de los demás niveles del *dataset*. El descenso leve de los resultados de este nivel para los nuevos valores fijados de dichos parámetros puede deberse a que se produzca alguna asociación cuyo valor de solapamiento con el *ground-truth* no sea de al menos un cincuenta por ciento.



## 7 Generalización de resultados

---

### 7.1 Introducción

En este capítulo se presentan los resultados finales obtenidos para cada una de las combinaciones entre detector y *tracker* utilizadas en este proyecto una vez añadidas todas las mejoras al módulo de *Matching* del sistema de seguimiento de múltiples personas (sección 7.2). Además, se han extraído una serie de conclusiones finales de los resultados obtenidos para cada configuración detector-*tracker* empleada (sección 7.3).

### 7.2 Resultados finales obtenidos

En esta sección se incluyen los resultados finales obtenidos para cada una de las configuraciones del detector-*tracker* utilizadas en este proyecto. El objetivo principal de este punto es observar los resultados que se obtienen al utilizar el detector Latent SVM y el *tracker* Filtro de Kalman una vez añadidas las mejoras que se introdujeron en el módulo de *Matching* del sistema de seguimiento de múltiples personas (véase sección 6).

#### 7.2.1 Resultados obtenidos utilizando el detector latent SVM y el *tracker* Filtro de Kalman

La Tabla 7-1 muestra los resultados medios obtenidos de las secuencias incluidas en cada uno de los niveles del *dataset* (explicado en la sección 3.4) para el detector Latent SVM y el *tracker* Filtro de Kalman una vez añadidas las mejoras introducidas en el módulo de *Matching* del sistema de seguimiento múltiple (véase sección 6). En la sección 6.2.4 se decidió fijar el parámetro  $\beta$  a 0.6 para la combinación del detector Latent SVM y el *tracker* Filtro de Partículas porque era el valor que obtenía mejores resultados globales para el *dataset* utilizado. Sin embargo, utilizándose el *tracker* Filtro de Kalman se ha fijado dicho valor a 0.4 para evitar problemas de pérdidas del objetivo por parte del *tracker* durante el proceso de seguimiento debidos al proceso de inicialización que necesita dicho *tracker* con las detecciones asociadas de *frames* posteriores. Los resultados obtenidos para cada una de las secuencias de cada uno de los niveles del *dataset* se han incluido en el Anexo 0 (subapartado B.6.1.3).

Como se puede apreciar, los resultados tanto de MOTA como de MOTP siguen descendiendo a medida que aumenta el nivel de complejidad de las secuencias evaluadas, como ocurría con los resultados obtenidos del sistema base (véase sección 5.5.1). Además, se sigue observando que el descenso de los resultados es más abrupto para MOTA que para MOTP porque como se comentó en la sección 5.5.1, MOTP medía la precisión entre los objetivos de salida del sistema base y las anotaciones del *ground-truth* para las asociaciones consideradas que superaban un solapamiento mayor del 50%, mientras que MOTA medía el rendimiento del sistema de seguimiento en función del número de falsos positivos, falsos negativos y cambios de identificador que desciende a medida que se incrementan las oclusiones y el número de objetivos en las secuencias.

Métrica	Resultados Nivel 1	Resultados Nivel 2	Resultados Nivel 3	Resultados Nivel 4	Resultados Nivel 5	Resultados Nivel 6
MOTP	0.73461	0.74896	0.70899	0.71449	0.71246	0.67948
MOTA	0.79943	0.68697	0.71274	0.68357	0.53581	0.60460
FP	0.10506	0.09631	0.02001	0.09079	0.14205	0.05487
TP	0.90449	0.78328	0.73275	0.77436	0.67785	0.65946
FN	0.09551	0.21672	0.26588	0.21544	0.31972	0.32969
ID switch	0.00	0.00	0.80	2.00	0.50	6.67

**Tabla 7-1: Resultados medios obtenidos para cada nivel del *dataset* utilizándose el *tracker* Filtro de Kalman tras las mejoras introducidas al sistema base.**

La Tabla 7-2 muestra una comparativa de los resultados medios obtenidos en cada uno de los niveles del *dataset* para el detector de personas Latent SVM y el *tracker* Filtro de Kalman tras introducirse las mejoras en el módulo de Matching del sistema de seguimiento múltiple (véase sección 6) con respecto al sistema base original. Además, muestra el porcentaje de mejora que se introduce tras dichas modificaciones. En negrita se han resaltado los resultados medios que obtienen mejores porcentajes de rendimiento, ya en el sistema base o después de las mejoras introducidas.

Como se puede apreciar, los resultados obtenidos para MOTP son parecidos a los resultados que se habían obtenido para el sistema base desarrollado originalmente. Esto se debe a lo que se ha comentado en puntos anteriores y es que MOTP mide la precisión de la localización del objetivo durante el proceso del seguimiento y depende únicamente de la precisión del detector utilizado en el sistema. En cambio, los resultados de MOTA aumentan considerablemente con respecto a los obtenidos para el sistema base, principalmente en los niveles más complejos del *dataset*. Esta mejora tan pronunciada en los niveles con mayor dificultad se debe principalmente a la reducción del parámetro  $\alpha$  (aprovechando de esta forma la gran precisión de los detectores de personas) y a la mejora introducida que permite inicializar el *tracker* una vez se ha detectado que ha perdido el objetivo durante su seguimiento. De esta forma, se reducen muchos falsos positivos (que deterioraban fuertemente los resultados objetivos del rendimiento) y se aumenta el número de verdaderos positivos.

Métrica	Nivel 1			Nivel 2		
	Resultados sist.base	Resultados tras mejoras	Mejora total sobre el sist.base	Resultados sist.base	Resultados tras mejoras	Mejora total sobre el sist.base
MOTP	<b>0.73578</b>	0.73461	-0.16%	0.73032	<b>0.74896</b>	2.55%
MOTA	0.78848	<b>0.79943</b>	1.39%	0.61482	<b>0.68697</b>	11.74%
FP	0.11054	0.10506	4.96%	0.13334	0.09631	27.77%
TP	0.89902	0.90449	0.61%	0.74819	0.78328	4.69%
FN	0.10099	0.09551	5.43%	0.25181	0.21672	13.94%
ID switch	0.00	0.00	0.00%	0.00	0.00	0.00%

Métrica	Nivel 3			Nivel 4		
	Resultados sist.base	Resultados tras mejoras	Mejora total sobre el sist.base	Resultados sist.base	Resultados tras mejoras	Mejora total sobre el sist.base
MOTP	<b>0.71326</b>	0.70899	-0.60%	0.70342	<b>0.71449</b>	1.57%
MOTA	0.67477	<b>0.71274</b>	5.63%	0.52726	<b>0.68357</b>	29.65%
FP	0.03293	0.02001	39.23%	0.21177	0.09079	57.13%
TP	0.70770	0.73275	3.54%	0.73902	0.77436	4.78%
FN	0.29093	0.26588	8.61%	0.24625	0.21544	12.51%
ID switch	0.80	0.80	0.00%	2.67	2.00	25.09%

Métrica	Nivel 5			Nivel 6		
	Resultados sist.base	Resultados tras mejoras	Mejora total sobre el sist.base	Resultados sist.base	Resultados tras mejoras	Mejora total sobre el sist.base
MOTP	0.69903	<b>0.71246</b>	1.92%	<b>0.68483</b>	0.67948	-0.78%
MOTA	0.32665	<b>0.53581</b>	64.03%	0.08979	<b>0.60460</b>	573.35%
FP	0.21743	0.14205	34.67%	0.29227	0.05487	81.23%
TP	0.54408	0.67785	24.59%	0.38202	0.65946	72.62%
FN	0.44737	0.31972	28.53%	0.59957	0.32969	45.01%
ID switch	3.00	0.50	83.33%	11.33	6.67	42.84%

**Tabla 7-2: Comparativa de resultados medios obtenidos para cada uno de los niveles del *dataset* entre el sistema base y el sistema final con las mejoras introducidas utilizándose el detector Latent SVM y el Filtro de Kalman. Resultados sistema base son los resultados obtenidos del sistema base original, Resultados tras mejoras hace referencia a los resultados finales obtenidos tras las mejoras introducidas y Mejora total sobre el sistema base es el porcentaje de mejora en los resultados finales con respecto a los resultados obtenidos del sistema base original, esto es, la acumulación de mejoras.**

## 7.2.2 Resultados obtenidos utilizando el detector latent SVM y el *tracker* Filtro de Partículas basado en color

La Tabla 7-3 muestra los resultados medios obtenidos de las secuencias incluidas en cada uno de los niveles del *dataset* (explicado en la sección 3.4) para el detector Latent SVM y el *tracker* Filtro de Partículas una vez añadidas las mejoras introducidas en el módulo de Matching del sistema de seguimiento múltiple (véase sección 6). Para añadir las mejoras al sistema base se utilizó esta configuración de detector-*tracker* por lo que los resultados que

se muestran coinciden con los que se obtuvieron tras la última mejora introducida (véase sección 6.2.6.2). Los resultados obtenidos para cada una de las secuencias del *dataset* se han incluido en el Anexo C.

Como se puede apreciar, el comportamiento es el mismo que el comentado en la sección 7.2.1 con un descenso de los resultados al incrementarse la complejidad de los niveles del *dataset*. Además, se puede observar también, el descenso más abrupto de los resultados para MOTA que para MOTP por el mismo motivo que se comentó en el punto anterior.

Métrica	Resultados Nivel 1	Resultados Nivel 2	Resultados Nivel 3	Resultados Nivel 4	Resultados Nivel 5	Resultados Nivel 6
MOTP	0.73223	0.75085	0.70743	0.69915	0.70312	0.70703
MOTA	0.82776	0.70802	0.72495	0.59860	0.57241	0.54441
FP	0.10559	0.09427	0.01842	0.12926	0.10904	0.10319
TP	0.93335	0.80230	0.74338	0.72785	0.68177	0.64760
FN	0.06665	0.19770	0.25525	0.26422	0.31549	0.33741
ID switch	0.00	0.00	0.80	1.67	1.50	9.00

**Tabla 7-3: Resultados medios obtenidos para cada nivel del *dataset* utilizándose el *tracker* Filtro de Partículas tras las mejoras introducidas al sistema base.**

La Tabla 7-4 muestra una comparativa de los resultados medios obtenidos en cada uno de los niveles del *dataset* para el detector de personas Latent SVM y el *tracker* Filtro de Partículas tras introducirse las mejoras en el módulo de Matching del sistema de seguimiento múltiple (véase sección 6) con respecto al sistema base original. Además, muestra el porcentaje de mejora que se introduce tras dichas modificaciones. En negrita se han resaltado los resultados medios que obtienen mejores porcentajes de rendimiento, ya en el sistema base o después de las mejoras introducidas.

Como se puede apreciar, el comportamiento es el mismo que el comentado en la sección 7.2.1 con un aumento más pronunciado en los resultados obtenidos para MOTA que en MOTP. La mejora tan pronunciada en los niveles con mayor dificultad es debida también a la reducción del parámetro  $\alpha$  (aprovechando de esta forma la gran precisión de los detectores de personas) y a la mejora introducida que permite inicializar el *tracker* una vez se ha detectado que ha perdido el objetivo durante su seguimiento. De esta forma, se reducen muchos falsos positivos (que deterioraban fuertemente los resultados objetivos del rendimiento) y se aumenta el número de verdaderos positivos.

Métrica	Nivel 1			Nivel 2		
	Resultados sist.base	Resultados tras mejoras	Mejora total sobre el sist.base	Resultados sist.base	Resultados tras mejoras	Mejora total sobre el sist.base
MOTP	0.72862	<b>0.73223</b>	0.50%	0.72540	<b>0.75085</b>	3.51%
MOTA	<b>0.83606</b>	0.82776	-0.99%	0.65629	<b>0.70802</b>	7.88%
FP	0.09911	0.10559	-6.54%	0.11495	0.09427	17.99%
TP	0.93517	0.93335	-0.19%	0.77138	0.80230	4.01%
FN	0.06483	0.06665	2.81%	0.22813	0.19770	13.34%
ID switch	0.00	0.00	0.00%	0.33	0.00	100.00%

Métrica	Nivel 3			Nivel 4		
	Resultados sist.base	Resultados tras mejoras	Mejora total sobre el sist.base	Resultados sist.base	Resultados tras mejoras	Mejora total sobre el sist.base
MOTP	0.70517	<b>0.70743</b>	0.32%	0.69751	<b>0.69915</b>	0.24%
MOTA	0.64020	<b>0.72495</b>	13.24%	0.47343	<b>0.59860</b>	26.44%
FP	0.04334	0.01842	57.50%	0.22339	0.12926	42.14%
TP	0.68357	0.74338	8.75%	0.69681	0.72785	4.45%
FN	0.31578	0.25525	19.17%	0.29414	0.26422	10.17%
ID switch	0.60	0.80	-33.33%	2.33	1.67	28.32%

Métrica	Nivel 5			Nivel 6		
	Resultados sist.base	Resultados tras mejoras	Mejora total sobre el sist.base	Resultados sist.base	Resultados tras mejoras	Mejora total sobre el sist.base
MOTP	0.66426	<b>0.70312</b>	5.85%	0.68225	<b>0.70703</b>	3.63%
MOTA	0.20504	<b>0.57241</b>	179.17%	0.23450	<b>0.54441</b>	132.16%
FP	0.27978	0.10904	61.03%	0.21602	0.10319	52.23%
TP	0.48481	0.68177	40.63%	0.45052	0.64760	43.75%
FN	0.51152	0.31549	38.32%	0.53416	0.33741	36.83%
ID switch	2.00	1.50	25.00%	9.33	9.00	3.53%

**Tabla 7-4: Comparativa de resultados medios obtenidos para cada uno de los niveles del *dataset* entre el sistema base y el sistema final con las mejoras introducidas utilizándose el detector Latent SVM y el Filtro de Partículas. Resultados sistema base son los resultados obtenidos del sistema base original, Resultados tras mejoras hace referencia a los resultados finales obtenidos tras las mejoras introducidas y Mejora total sobre el sistema base es el porcentaje de mejora en los resultados finales con respecto a los resultados obtenidos del sistema base original, esto es, la acumulación de mejoras.**

### 7.2.3 Comparativa de resultados finales

La Tabla 7-5 muestra una comparativa de los resultados medios obtenidos para cada uno de los niveles del *dataset* para el detector Latent SVM y los *trackers* Filtro de Kalman y Filtro de Partículas tras introducirse las mejoras en el módulo de Matching del sistema de seguimiento múltiple (véase sección 6) con respecto al sistema base original. En negrita se ha resaltado la combinación detector-*tracker* que obtiene mejores resultados.

Como se puede apreciar, los resultados para MOTP son prácticamente los mismos para las dos combinaciones de detector-*tracker*, con diferencias máximas de un 1% aproximadamente, a excepción del nivel 6 donde el Filtro de Partículas supera al Filtro de Kalman en un 2.76%. Esto era de esperar ya que la precisión del sistema de seguimiento de múltiples personas depende en gran medida de la precisión del *bounding box* de la salida del detector de personas utilizado, que en las dos combinaciones es el mismo.

Por el contrario, si observamos los resultados obtenidos para MOTA se observa que en la mayoría de los niveles la combinación del detector Latent SVM y el *tracker* Filtro de Partículas obtiene unos resultados ligeramente mejores. Sin embargo, en los niveles 4 y 6 del *dataset* utilizado el Filtro de Kalman obtiene mejores resultados. En el nivel 4 obtiene una mejora de resultados medios de un 8.50%, mientras que en el nivel 6 de un 6.02%. Esta mejora tan pronunciada en el nivel 4 se debe a que en una de las secuencias empleadas se produce una oclusión total que el Filtro de Partículas no es capaz de soportar y que genera, por tanto, un mayor número de falsos positivos y un menor número de verdaderos positivos (que degradan notablemente los resultados objetivos del rendimiento del sistema). Por otro lado, la mejora producida del Filtro de Kalman con respecto al Filtro de Partículas en el nivel 6 se debe a que las secuencias empleadas en este nivel poseen un movimiento rápido de los objetivos en la escena que propicia que este *tracker* pueda perder al objetivo más fácilmente para los parámetros establecidos (ver sección 5.4.2.2). Si se aumenta el valor fijado de los parámetros  $\sigma_x$  y  $\sigma_y$  (varianzas de la posición del centroide del objetivo  $x, y$ ) los resultados mejoran para esta combinación. Además, esta mejora de resultados en este nivel para el *tracker* Filtro de Kalman puede deberse también a que en estas secuencias se producen un número alto de oclusiones que este *tracker* las atraviesa prediciendo linealmente la posición del objetivo en instantes posteriores dando lugar a un número mayor de verdaderos positivos si la predicción es buena, mientras que el Filtro de Partículas se recupera de la oclusión una vez desaparece dando lugar a un mayor número de falsos positivos y a un menor número de verdaderos positivos.

Métrica	Nivel 1			Nivel 2			Nivel 3		
	KALMAN	PARTICLE	Diferencia resultados	KALMAN	PARTICLE	Diferencia resultados	KALMAN	PARTICLE	Diferencia resultados
MOTP	<b>0.73461</b>	0.73223	-0.24%	0.74896	<b>0.75085</b>	0.19%	<b>0.70899</b>	0.70743	-0.16%
MOTA	0.79943	<b>0.82776</b>	2.83%	0.68697	<b>0.70802</b>	2.11%	0.71274	<b>0.72495</b>	1.22%
FP	0.10506	0.10559	0.05%	0.09631	0.09427	0.20%	0.02001	0.01842	0.16%
TP	0.90449	0.93335	2.89%	0.78328	0.80230	1.90%	0.73275	0.74338	1.06%
FN	0.09551	0.06665	-2.89%	0.21672	0.19770	1.90%	0.26588	0.25525	1.06%
ID switch	0.00	0.00	0.00	0.00	0.00	0.00	0.80	0.80	0.00

Métrica	Nivel 4			Nivel 5			Nivel 6		
	KALMAN	PARTICLE	Diferencia resultados	KALMAN	PARTICLE	Diferencia resultados	KALMAN	PARTICLE	Diferencia resultados
MOTP	<b>0.71449</b>	0.69915	1.53%	<b>0.71246</b>	0.70312	-0.93%	0.67948	<b>0.70703</b>	-2.76%
MOTA	<b>0.68357</b>	0.59860	8.50%	0.53581	<b>0.57241</b>	3.66%	<b>0.60460</b>	0.54441	6.02%
FP	0.09079	0.12926	3.85%	0.14205	0.10904	3.30%	0.05487	0.10319	4.83%
TP	0.77436	0.72785	4.65%	0.67785	0.68177	0.39%	0.65946	0.64760	1.19%
FN	0.21544	0.26422	4.88%	0.31972	0.31549	0.42%	0.32969	0.33741	0.77%
ID switch	2.00	1.67	-0.33	0.50	1.50	-1.00	6.67	9.00	2.33

**Tabla 7-5: Comparativa de resultados medios obtenidos para cada nivel del *dataset* utilizando el detector Latent SVM y los *trackers* Filtro de Kalman y Filtro de Partículas con las mejoras introducidas. La primera columna de cada nivel muestra los resultados obtenidos utilizando el *tracker* Filtro de Kalman, la segunda columna indica los resultados obtenidos utilizando el *tracker* Filtro de Partículas y la tercera columna indica la diferencia en porcentaje entre la mejor y peor combinación de detector-*tracker* en cuanto a los resultados obtenidos para la medida de rendimiento MOTA.**

### 7.3 Conclusiones

Durante este capítulo se han presentado los resultados finales obtenidos para cada una de las combinaciones entre detector y *tracker* utilizadas en este proyecto una vez añadidas todas las mejoras al módulo de Matching del sistema de seguimiento de múltiples personas (sección 7.2).

En la sección 7.2.1 se han detallado los resultados finales obtenidos para la combinación del detector Latent SVM y el *tracker* Filtro de Kalman. Observándose los resultados obtenidos se ha apreciado que los resultados para MOTP son parecidos a los resultados que se habían obtenido para el sistema base desarrollado originalmente. Esto era de esperar ya que MOTP mide la precisión de la localización del objetivo durante el proceso del seguimiento y depende únicamente de la precisión del detector utilizado en el sistema. En cambio, los resultados para MOTA han aumentado considerablemente con respecto a los obtenidos para el sistema base, principalmente en los niveles más complejos del *dataset* donde las secuencias incluidas contienen un elevado número de personas y un movimiento rápido de los objetivos. En este tipo de secuencias es más sencillo que el Filtro de Kalman pueda perder al objetivo durante el proceso de seguimiento si el detector de personas produce muchos falsos negativos (*missing detections*), y por tanto, no se lleva a cabo su proceso de corrección o actualización de la predicción realizada. Por ello, esta mejora tan pronunciada en los niveles con mayor dificultad se debía principalmente a la reducción del parámetro  $\alpha$  (aprovechando de esta forma la gran precisión de los detectores de personas) y

a la mejora introducida que permitía inicializar el *tracker* una vez se había detectado que había perdido el objetivo durante su seguimiento.

En la sección 7.2.2 se han incluido los resultados finales obtenidos para la combinación del detector Latent SVM y el *tracker* Filtro de Partículas. Observándose los resultados obtenidos, se ha apreciado el mismo comportamiento que el comentado en el párrafo anterior con un aumento más pronunciado en los resultados obtenidos para MOTA que en MOTP. La mejora tan pronunciada en los niveles con mayor dificultad que como se ha comentado anteriormente poseen un elevado número de personas y un movimiento rápido de los objetivos se debía también a la reducción del parámetro  $\alpha$  y a la mejora introducida que permitía inicializar el *tracker* una vez se había detectado que había perdido el objetivo durante su seguimiento. Sin embargo, también se ha demostrado que aumentando las varianzas de la posición del centroide del objetivo en este tipo de secuencias con un movimiento más rápido de los objetivos se reducían las pérdidas de los objetivos por parte de este *tracker* durante el proceso de seguimiento. Por tanto, este *tracker* permite adaptarse mejor a este tipo de problemas mediante un ajuste correcto de sus parámetros, siendo de este modo, menos dependiente del correcto funcionamiento del detector empleado en el sistema.

## 8 Conclusiones y trabajo futuro

---

### 8.1 Conclusiones

El objetivo principal de este proyecto era el diseño e implementación de un algoritmo de seguimiento de múltiples personas basado en la combinación de la información proporcionada por un detector de personas y un *tracker*.

Para ello, en primer lugar, se realizó un estudio detallado del estado del arte del seguimiento de múltiples personas en vídeo basado en detección, analizando los algoritmos de detección de personas y de seguimiento de objetos existentes en la literatura actual, así como, los algoritmos de seguimiento que combinan el seguimiento con la detección. De este estudio se extrajeron las principales características y limitaciones que presentaba cada uno de los enfoques analizados.

Seguidamente, se desarrolló un protocolo de evaluación que era necesario para evaluar el algoritmo de seguimiento múltiple que tenía que ser implementado. Para este fin, se construyó un *dataset* con diferentes niveles de complejidad y se seleccionaron algunas de las métricas más utilizadas por la comunidad investigadora para medir el rendimiento del sistema de seguimiento. Por tanto, se propuso una metodología de evaluación objetiva con el fin de obtener unos resultados objetivos del algoritmo que iba a ser desarrollado y una metodología de evaluación subjetiva mediante la cual se obtenía una salida visual del algoritmo de seguimiento.

Una vez desarrollado el protocolo de evaluación, se seleccionaron una serie de algoritmos de seguimiento y de detección del estado del arte y se analizó el funcionamiento de cada uno de ellos.

Tras este análisis, se desarrolló un sistema base para el seguimiento de múltiples personas basado en detección. Este sistema incluía un módulo de asociación de identidades mediante el cual se combinaba la información proporcionada por un *tracker* y por un algoritmo de detección y manejaba el nacimiento de nuevos objetivos y la muerte de objetivos que llevaban cierto tiempo sin obtener ninguna asociación. Se obtuvieron los resultados objetivos para cada una de las combinaciones detector-*tracker* utilizadas en el proyecto y se seleccionó aquella que obtenía mejores resultados globales para añadir las mejoras al módulo de asociación de identidades del sistema base.

Después de desarrollar el sistema base se introdujeron una serie de mejoras a su módulo de asociación de identidades para mejorar el comportamiento de dicho sistema. Además, para demostrar que realmente el rendimiento del sistema mejoraba se obtuvieron los resultados objetivos tras la inserción de cada una de las modificaciones y se compararon con respecto a los que se habían obtenido para el sistema original desarrollado anteriormente.

Por último, una vez introducidas todas las mejoras al sistema base se obtuvieron los resultados finales para cada una de las combinaciones detector-*tracker* utilizadas en el proyecto, para confirmar que el rendimiento del sistema incrementaba también con respecto a los resultados obtenidos del sistema base.

Por tanto, se puede concluir que una vez desarrollado el sistema y obtenido los resultados del rendimiento del algoritmo se han conseguido alcanzar los objetivos que se planteaban antes del comienzo del proyecto realizado.

## **8.2 Trabajo futuro**

Tras la realización de este proyecto, son numerosas las opciones de trabajo futuro en las que se considera que se podría avanzar a partir de este trabajo con el fin de mejorar los resultados obtenidos. A continuación, se resumen algunas ideas que se podrían utilizar.

En primer lugar, se podría continuar mejorando el módulo de *Matching* del algoritmo implementado con el objetivo de aumentar el número de verdaderos positivos y de reducir el número de falsos positivos y el número de cambios de identidad entre objetivos. Para ello, algunas mejoras que se podrían introducir serían la utilización de la información de la cámara, de forma que se pudiesen eliminar detecciones espurias que producen un deterioro de los resultados, o la modificación del modelo de características empleado en el proceso de asociación de identidades.

También, se podrían probar diferentes combinaciones de detector-*tracker* existentes en la literatura actual con la intención de obtener mejores resultados que los obtenidos con las combinaciones utilizadas en este proyecto. Además, se podría estudiar fusionar la información proporcionada por varios detectores y *trackers*. De esta forma, se podrían paliar defectos de su funcionamiento en determinadas situaciones, por ejemplo, en casos en los que un detector proporcione la información de detección de un objetivo, mientras que otros obtengan para ese mismo objetivo un falso negativo.

Otra opción, sería mejorar o diseñar nuevos algoritmos de detección de personas o de seguimiento de objetos, con la finalidad de obtener tasas más bajas de falsos negativos y aumentar la robustez a las oclusiones.

Además, el sistema implementado se podría ampliar a sistemas multicámaras o a sistemas con cámaras en movimiento ya que este proyecto está orientado a la implementación de un algoritmo en un entorno de cámaras fijas.

Por otro lado, se podría mejorar la eficiencia del sistema implementado, de tal forma que se mejorasen aspectos técnicos del código y se incluyese un procesado en paralelo de la ejecución del detector de personas y del *tracker* que son dos procesos independientes pero que se realizan secuencialmente.

# Referencias

---

- [1] Maggio, Emilio, and Cavallaro, Andrea. Video tracking: theory and practice. John Wiley & Sons, 2011.
- [2] Yang, Hanxuan, et al. "Recent advances and trends in visual tracking: A review." *Neurocomputing*, 74(18):3823-3831, 2011.
- [3] Andriluka, Mykhaylo, Roth, Stefan, and Schiele, Bernt. "People-tracking-by-detection and people-detection-by-tracking." In *Proc. of CVPR 2008*, 1-8.
- [4] Breitenstein, Michael D., et al. "Online multiperson tracking-by-detection from a single, uncalibrated camera." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(9):1820-1833, 2011.
- [5] Enzweiler, Markus, and Gavrilu, Dariu M. "Monocular pedestrian detection: Survey and experiments." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(12): 2179-2195, 2009.
- [6] Geronimo, David, et al. "Survey of pedestrian detection for advanced driver assistance systems." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(7): 1239-1258, 2010.
- [7] Dollar, Piotr, et al. "Pedestrian detection: An evaluation of the state of the art." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(4): 743-761, 2012.
- [8] Fortmann, Thomas E., Bar-Shalom, Yaakov, and Scheffe, Molly. "Sonar tracking of multiple targets using joint probabilistic data association." *IEEE Journal of Oceanic Engineering*, 8(3): 173-184, 1983.
- [9] Reid, Donald B. "An algorithm for tracking multiple targets." *IEEE Transactions on Automatic Control*, 24(6): 843-854, 1979.
- [10] Kuhn, Harold W. "The Hungarian method for the assignment problem." *Naval research logistics quarterly*, 2(1-2): 83-97, 1955.
- [11] Wu, Bo, and Nevatia, Ram. "Detection and tracking of multiple, partially occluded humans by bayesian combination of edgelet based part detectors." *International Journal of Computer Vision* 75(2): 247-266, 2007.
- [12] Hu, Weiming, et al. "A survey on visual surveillance of object motion and behaviors." *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 34(3): 334-352, 2004.
- [13] Valera, M., and Velastin, S. A. "Intelligent distributed surveillance systems: a review." *IEE Proceedings - Vision, Image and Signal Processing*, 2(152):192-204, 2005.
- [14] Martín, Alvaro García. Contributions to robust people detection in video-surveillance. PhD Diss. Universidad Autónoma de Madrid, 2013.
- [15] Cutler, Ross, and Davis, Larry S. "Robust real-time periodic motion detection, analysis, and applications." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8): 781-796, 2000.
- [16] Haritaoglu, Ismail, Harwood, David, and Davis, Larry S. "W 4: Real-time surveillance of people and their activities." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8): 809-830, 2000.
- [17] Zhao, Tao, and Nevatia, Ramakant. "Tracking multiple humans in complex situations." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9): 1208-1221, 2004.
- [18] Zhou, Jianpeng, and Hoang, Jack. "Real time robust human detection and tracking system." In *Proc. of CVPR 2005*, 149.
- [19] Hussein, Mohamed, et al. "Real-time human detection, tracking, and verification in uncontrolled camera motion environments." In *Proc. of ICVS 2006*, 41.
- [20] Fernández-Carbajales, Víctor, García, Miguel Ángel and Martínez José M. "Robust People Detection by Fusion of Evidence from Multiple Methods." In *Proc. of WIAMIS 2008*, 55-58.

- [21] Kilambi, Prahlad, et al. "Estimating pedestrian counts in groups." *Computer Vision and Image Understanding*, 110(1): 43-59, 2008.
- [22] Sprague, Nathan, and Luo, Jiebo. "Clothed people detection in still images." In *Proc. of CVPR 2002*, 585-589.
- [23] Harasse, Sebastien, Bonnaud, Laurent, and Michel Desvignes. "Human model for people detection in dynamic scenes." In *Proc. of ICPR 2006*, 335-354.
- [24] Giebel, Jan, Gavrilu, Darin M., and Schnörr, Christoph. "A bayesian framework for multi-cue 3d object tracking." In *Proc. of ECCV 2004*, 241-252.
- [25] Alonso, Ignacio Parra, et al. "Combination of feature extraction methods for SVM pedestrian detection." *IEEE Transactions on Intelligent Transportation Systems*, 8(2): 292-307, 2007.
- [26] Gavrilu, Dariu M., and Munder, Stefan. "Multi-cue pedestrian detection and tracking from a moving vehicle." *International journal of computer vision*, 73(1): 41-59, 2007.
- [27] Xu, Fengliang, and Fujimura, Kikuo. "Human detection using depth and gray images." In *Proc. of AVSS 2003*, 115-121.
- [28] Koenig, Nathan. "Toward real-time human detection and tracking in diverse environments." In *Proc. of ICDL 2007*, 94-98.
- [29] Viola, Paul, Jones, Michael J., and Snow, Daniel. "Detecting pedestrians using patterns of motion and appearance." In *Proc. of ICCV 2003*, 734-741.
- [30] Okuma, Kenji, et al. "A boosted particle filter: Multitarget detection and tracking." In *Proc. of ECCV 2004*, 28-39.
- [31] Sidenbladh, Hedvig. "Detecting human motion with support vector machines." In *Proc. of ICPR 2004*, 188-191.
- [32] Viola, Paul, and Jones, Michael J. "Robust real-time face detection." *International journal of computer vision*, 57(2): 137-154, 2004.
- [33] Felzenszwalb, Pedro F., et al. "Object detection with discriminatively trained part-based models." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9): 1627-1645, 2010.
- [34] Dalal, Navneet, and Triggs, Bill. "Histograms of oriented gradients for human detection." In *Proc. of CVPR 2005*, 886-893.
- [35] Leibe, Bastian, and, Schiele, Bernt. "Scale-invariant object categorization using a scale-adaptive mean-shift search." In *Proc. of DAGM 2004*, 145-153.
- [36] Seemann, Edgar, and, Schiele, Bernt. "Cross-articulation learning for robust detection of pedestrians." In *Proc. of DAGM 2006*, 242-252.
- [37] Leibe, Bastian, Schindler, Konrad, and Van Gool, Luc. "Coupled detection and trajectory estimation for multi-object tracking." In *Proc. of ICCV 2007*, 1-8.
- [38] Andriluka, Mykhaylo, Roth, Stefan, and Schiele, Bernt. "Pictorial structures revisited: People detection and articulated pose estimation." In *Proc. of CVPR 2009*, 1014-1021.
- [39] Leibe, Bastian, Leonardis, Aleš, and Schiele, Bernt. "Robust object detection with interleaved categorization and segmentation." *International journal of computer vision*, 77(1-3): 259-289, 2008.
- [40] Cutler, Ross, and Davis, Larry S. "Robust real-time periodic motion detection, analysis, and applications." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8): 781-796, 2000.
- [41] Cui, Xinyi, Liu, Yazhou, and Shan, Shiguang. "3d haar-like features for pedestrian detection." In *Proc. of ICME 2007*, 1263-1266.
- [42] Yilmaz, Alper, Javed, Omar, and Shah, Mubarak. "Object tracking: A survey." *Acm computing surveys (CSUR)*, 38(4): 13, 2006.
- [43] Cannons, Kevin. "A review of visual tracking." *Technical Report CSE*, 2008.
- [44] Park, Jung Me, and Murphey, Yi Lu. "Edge detection in grayscale, color, and range images." *Wiley Encyclopedia of Computer Science and Engineering*, 2008.
- [45] Canny, John. "A computational approach to edge detection." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(8): 679-698, 1986.

- [46] Marr, David, and Hildreth, Ellen. "Theory of edge detection." *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 207(1167): 187-217, 1980.
- [47] Lowe, David G. "Distinctive image features from scale-invariant keypoints." *International journal of computer vision* 60(2): 91-110, 2004.
- [48] Bay, Herbert, Tuytelaars, Tinne and Van Gool, Luc. "Surf: Speeded up robust features." In *Proc. of ECCV 2006*, 404-417.
- [49] Cheung, Sen-Ching S., and Kamath, Chandrika. "Robust background subtraction with foreground validation for urban traffic video." *EURASIP Journal on Advances in Signal Processing*, 2005(14): 2330-2340, 1900.
- [50] Donatello, Conte, et al. "An experimental evaluation of foreground detection algorithms in real scenes." *EURASIP Journal on Advances in Signal Processing*, 7(2010): 1-10, 2010.
- [51] Khan, Ishtiaq Rasool, and Farbiz, Farzam. "A back projection scheme for accurate mean shift based tracking." In *Proc. of ICIP 2010*, 33-36.
- [52] Kalman, Rudolph Emil. "A new approach to linear filtering and prediction problems." *Journal of basic Engineering*, 82(1): 35-45, 1960.
- [53] Gordon, Neil J., Salmond, David J., and Smith, Adrian FM. "Novel approach to nonlinear/non-Gaussian Bayesian state estimation." *IEE Proceedings F (Radar and Signal Processing)*, 2(140), 1993.
- [54] Kalal, Zdenek, Mikolajczyk, Krystian, and Matas, Jiri. "Forward-backward error: Automatic detection of tracking failures." In *Proc. of ICPR 2010*, 2756-2759.
- [55] Kim, Changick, and Hwang, Jenq-Neng. "Fast and automatic video object segmentation and tracking for content-based applications." *IEEE Transactions on Circuits and Systems for Video Technology*, 12(2): 122-129, 2002.
- [56] Yilmaz, Alper, Li, Xin, and Shah, Mubarak. "Object contour tracking using level sets." In *Proc. of ACCV 2004*, 432-437.
- [57] Chen, Yunqiang, Rui, Yong, and Huang, Thomas S. "JPDAF based HMM for real-time contour tracking." In *Proc. of CVPR 2001*, 543-550.
- [58] Maggio, Emilio, and Cavallaro, Andrea. "Hybrid particle filter and mean shift tracker with adaptive transition model." In *Proc. of ICASSP 2005*, 221-224.
- [59] Stauffer, Chris, and Grimson, W. Eric L. "Learning patterns of activity using real-time tracking." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8): 747-757, 2000.
- [60] Brunelli, Roberto. *Template matching techniques in computer vision: theory and practice*. John Wiley & Sons, 2009.
- [61] Zhong, Yu, Jain, Anil K., and Dubuisson-Jolly, M-P. "Object tracking using deformable templates." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(5): 544-549, 2000.
- [62] Berclaz, Jerome, Fleuret, Francois and Fua, Pascal. "Robust people tracking with global trajectory optimization." In *Proc. of CVPR 2006*, 744-750.
- [63] Leibe, Bastian, et al. "Coupled object detection and tracking from static cameras and moving vehicles." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(10): 1683-1698, 2008.
- [64] Li, Yuan, Huang, Chang, and Nevatia, Ram. "Learning to associate: Hybridboosted multi-target tracker for crowded scene." In *Proc. of CVPR 2009*, 2953-2960.
- [65] Zhao, Tao, and Nevatia, Ramakant. "Tracking multiple humans in complex situations." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9): 1208-1221, 2004.
- [66] Khan, Saad M., and Shah, Mubarak. "A multiview approach to tracking people in crowded scenes using a planar homography constraint." In *Proc. of ECCV 2006*, 133-146.
- [67] Lanz, Oswald. "Approximate bayesian multibody tracking." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(9): 1436-1449, 2006.
- [68] Song, Xuan, et al. "Vision-based multiple interacting targets tracking via on-line supervised learning." In *Proc. of ECCV 2008*, 642-655.

- [69] Huang, Chang, Bo Wu, and Nevatia, Ramakant. "Robust object tracking by hierarchical association of detection responses." In Proc. of ECCV 2008, 788-801.
- [70] Perera, AG Amitha, et al. "Multi-object tracking through simultaneous long occlusions and split-merge conditions." In Proc. of CVPR 2006, 666-673.
- [71] Zhang, Li, Li, Yuan, and Nevatia, Ramakant. "Global data association for multi-object tracking using network flows." In Proc. of CVPR 2008, 1-8.
- [72] Vermaak, Jaco, Doucet, Arnaud, and Pérez, Patrick. "Maintaining multimodality through mixture tracking." In Proc. of ICCV 2003, 1110-1116.
- [73] Cai, Yizheng, Freitas, Nando de, and Little, James J. "Robust visual tracking for multiple targets". In Proc. of ECCV 2006, 107-118.
- [74] Ess, Andreas, et al. "Robust multiperson tracking from a mobile platform." IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(10): 1831-1846, 2009.
- [75] Pellegrini, Stefano, et al. "You'll never walk alone: Modeling social behavior for multi-target tracking." In Proc. of ICCV 2009, 261-268.
- [76] Avidan, Shai. "Ensemble tracking." IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(2): 261-271, 2007.
- [77] Grabner, Helmut, and Bischof, Horst. "On-line boosting and vision." In Proc. of CVPR 2006, 260-267.
- [78] Li, Yuan, et al. "Tracking in low frame rate video: A cascade particle filter with discriminative observers of different life spans." IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(10): 1728-1740, 2008.
- [79] Wu, Bo, et al. "Robust object tracking based on detection with soft decision." In Proc. of WMVC 2008, 1-8.
- [80] Verma, Ragini Choudhury, Schmid, Cordelia and Mikolajczyk, Krystian. "Face detection and tracking in a video by propagating detection probabilities." IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(10): 1215-1228, 2003.
- [81] Ren, Xiaofeng. "Finding people in archive films through tracking." In Proc. of CVPR 2008, 1-8.
- [82] Nawaz, Tahir, Poiesi, Fabio and Cavallaro, Andrea. "Measures of effective video tracking." IEEE Transactions on Image Processing, 1(3):376-388, 2014.
- [83] Black, James, Ellis, Tim and Rosin, Paul. "A novel method for video tracking performance evaluation." In Proc. of PETS 2003, 125-132.
- [84] Hoffman, John R., and Mahler, Ronald PS. "Multitarget miss distance via optimal assignment." IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, 34(3): 327-336.
- [85] Munkres, James. "Algorithms for the assignment and transportation problems." Journal of the Society for Industrial & Applied Mathematics, 5(1): 32-38, 1957.
- [86] Brown, L. M., Senior, A. W., Tian, Y.-L., Connell, J., Hampapur, A., Shu, C. f., Merkl, H., and Lu, M. "Performance evaluation of surveillance systems under varying conditions," In Proc. of WPETS 2005, pp. 1-8.
- [87] Yin, Fei, Makris, Dimitrios and Velastin, Sergio A. "Performance evaluation of object tracking algorithms." In Proc. of PETS 2007, 17-24.
- [88] Keni, Bernardin, and Rainer, Stiefelhagen. "Evaluating multiple object tracking performance: the CLEAR MOT metrics." EURASIP Journal on Image and Video Processing, 1(2008):1-10, 2008.
- [89] Andriyenko, Anton, Schindler, Konrad, and Roth, Stefan. "Discrete-continuous optimization for multi-target tracking." In Proc. of CVPR 2012, 1926-1933.
- [90] Berclaz, Jerome, et al. "Multiple object tracking using k-shortest paths optimization." IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(9): 1806-1819, 2011.
- [91] Breitenstein, Michael D., et al. "Robust tracking-by-detection using a detector confidence particle filter." In Proc. of ICCV 2009, 1515-1522.
- [92] Andriyenko, Anton, and Schindler, Konrad. "Multi-target tracking by continuous energy minimization." In Proc. of CVPR 2011, 1265-1272.

- [93] Zhang, Jianming, Presti, Liliana Lo, and Sclaroff, Stan. "Online multi-person tracking by tracker hierarchy." In Proc. of AVSS 2012, 379-385.
- [94] Ramírez, Álvaro Solera. "El Filtro de Kalman." Documento de trabajo del Banco Central de Costa Rica, elaborado en la División Económica, Departamento de Investigaciones Económicas, 2003.
- [95] Arulampalam, M. Sanjeev, et al. "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking." IEEE Transactions on Signal Processing, 50(2): 174-188, 2002.
- [96] Nummiaro, Katja, Koller-Meier, Esther, and Van Gool, Luc. "An adaptive color-based particle filter." Image and vision computing, 21(1): 99-110, 2003.



# Anexos

## A. Resultados obtenidos del sistema base

### A.1. Parámetros y resultados obtenidos utilizando el detector latent SVM y el tracker Filtro de Kalman para cada nivel del dataset

#### A.1.1 Nivel 1

Parámetros	Secuencia 1	Secuencia 2	Secuencia 3
UMBRAL_H	40	40	40
UMBRAL_W	40	40	40
UMBRAL_FEATURE	40	40	30
RADIO	40	40	40

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.78007	0.72907	0.69821
MOTA	0.86473	0.66554	0.83516
FP	0.02657	0.27365	0.03140
TP	0.89130	0.93919	0.86656
FN	0.10870	0.06081	0.13344
ID switch	0	0	0

#### A.1.2 Nivel 2

Parámetros	Secuencia 1	Secuencia 2	Secuencia 3
UMBRAL_H	60	40	80
UMBRAL_W	60	40	80
UMBRAL_FEATURE	50	70	80
RADIO	40	40	60

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.76258	0.66489	0.76349
MOTA	0.59178	0.53676	0.71591
FP	0.25404	0.12500	0.02110
TP	0.84581	0.66176	0.73701
FN	0.15419	0.33824	0.26299
ID switch	0	0	0

#### A.1.3 Nivel 3

Parámetros	Secuencia 1	Secuencia 2	Secuencia 3	Secuencia 4	Secuencia 5
UMBRAL_H	90	50	70	70	70
UMBRAL_W	90	50	70	70	70
UMBRAL_FEATURE	80	70	50	50	60
RADIO	80	60	40	40	40

Métrica	Secuencia 1	Secuencia 2	Secuencia 3	Secuencia 4	Secuencia 5
MOTP	0.70846	0.68499	0.71537	0.73281	0.72467
MOTA	0.73985	0.57588	0.80101	0.85034	0.40678
FP	0.03158	0.00681	0.00000	0.00340	0.12288
TP	0.77143	0.58268	0.80101	0.85374	0.52966
FN	0.22707	0.41537	0.19899	0.14286	0.47034
ID switch	1	2	0	1	0

#### A.1.4 Nivel 4

Parámetros	Secuencia 1	Secuencia 2	Secuencia 3
UMBRAL_H	40	70	60
UMBRAL_W	40	70	60
UMBRAL_FEATURE	50	80	80
RADIO	40	80	60

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.69095	0.72592	0.69339
MOTA	0.74265	0.64865	0.19048
FP	0.06128	0.11824	0.45578
TP	0.80392	0.76689	0.64626
FN	0.19608	0.22973	0.31293
ID switch	0	2	6

### A.1.5 Nivel 5

Parámetros	Secuencia 1	Secuencia 2
UMBRAL_H	60	80
UMBRAL_W	60	80
UMBRAL_FEATURE	120	60
RADIO	70	80

Métrica	Secuencia 1	Secuencia 2
MOTP	0.69244	0.70561
MOTA	0.37073	0.28257
FP	0.20000	0.23486
TP	0.57073	0.51743
FN	0.41951	0.47523
ID switch	2	4

### A.1.6 Nivel 6

Parámetros	Secuencia 1	Secuencia 2	Secuencia 3
UMBRAL_H	50	40	40
UMBRAL_W	50	40	40
UMBRAL_FEATURE	80	60	60
RADIO	70	50	40

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.67262	0.68326	0.69862
MOTA	-0.06195	0.15798	0.17333
FP	0.41239	0.29244	0.17185
TP	0.35044	0.45042	0.34519
FN	0.63186	0.53277	0.63407
ID switch	10	10	14

## A.2. Parámetros y resultados obtenidos utilizando el detector latent SVM y el tracker Filtro de Partículas para cada nivel del dataset

### A.2.1 Nivel 1

Parámetros	Secuencia 1	Secuencia 2	Secuencia 3
UMBRAL_H	40	40	90
UMBRAL_W	40	40	90
UMBRAL_FEATURE	40	40	70
RADIO	30	40	40

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.77164	0.72420	0.69002
MOTA	0.92271	0.66554	0.91994
FP	0.00483	0.27365	0.01884
TP	0.92754	0.93919	0.93878
FN	0.07246	0.06081	0.06122
ID switch	0	0	0

### A.2.2 Nivel 2

Parámetros	Secuencia 1	Secuencia 2	Secuencia 3
UMBRAL_H	60	40	80
UMBRAL_W	60	40	80
UMBRAL_FEATURE	50	70	60
RADIO	40	40	60

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.77245	0.64643	0.75732
MOTA	0.60352	0.63971	0.72565
FP	0.2551	0.073529	0.01623
TP	0.85903	0.71324	0.74188
FN	0.1395	0.28676	0.25812
ID switch	1	0	0

### A.2.3 Nivel 3

Parámetros	Secuencia 1	Secuencia 2	Secuencia 3	Secuencia 4	Secuencia 5
UMBRAL_H	90	40	70	50	70
UMBRAL_W	90	40	70	50	70
UMBRAL_FEATURE	80	70	50	50	60
RADIO	80	60	40	40	40

Métrica	Secuencia 1	Secuencia 2	Secuencia 3	Secuencia 4	Secuencia 5
MOTP	0.70160	0.69447	0.70624	0.70724	0.71630
MOTA	0.74286	0.58755	0.80101	0.81108	0.25847
FP	0.03158	0.00292	0.00000	0.00000	0.18220
TP	0.77444	0.59047	0.80101	0.81108	0.44068
FN	0.22406	0.40759	0.19899	0.18892	0.55932
ID switch	1	2	0	0	0

## A.2.4 Nivel 4

Parámetros	Secuencia 1	Secuencia 2	Secuencia 3
UMBRAL_H	40	70	60
UMBRAL_W	40	70	60
UMBRAL_FEATURE	50	80	110
RADIO	50	80	60

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.69876	0.70852	0.68525
MOTA	0.7451	0.49831	0.17687
FP	0.06618	0.21622	0.38776
TP	0.81127	0.71453	0.56463
FN	0.18873	0.27872	0.41497
ID switch	0	4	3

## A.2.5 Nivel 5

Parámetros	Secuencia 1	Secuencia 2
UMBRAL_H	60	80
UMBRAL_W	60	80
UMBRAL_FEATURE	70	60
RADIO	40	80

Métrica	Secuencia 1	Secuencia 2
MOTP	0.64847	0.68005
MOTA	0.30732	0.10275
FP	0.22927	0.33028
TP	0.53659	0.43303
FN	0.46341	0.55963
ID switch	0	4

## A.2.6 Nivel 6

Parámetros	Secuencia 1	Secuencia 2	Secuencia 3
UMBRAL_H	40	70	70
UMBRAL_W	40	70	70
UMBRAL_FEATURE	60	80	80
RADIO	40	90	70

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.69027	0.676	0.68047
MOTA	0.33982	0.21849	0.14519
FP	0.18053	0.28235	0.18519
TP	0.52035	0.50084	0.33037
FN	0.46195	0.48571	0.65481
ID switch	10	8	10



## B. Resultados obtenidos para cada una de las modificaciones del sistema base

### B.1. Primera modificación

#### B.1.1 Resultados obtenidos para cada nivel del dataset

##### B.1.1.1 Nivel 1

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.78435	0.72724	0.69564
MOTA	0.91304	0.66554	0.91994
FP	0.00242	0.27365	0.01570
TP	0.91546	0.93919	0.93564
FN	0.08451	0.06081	0.06436
ID switch	0	0	0

##### B.1.1.2 Nivel 2

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.77336	0.68066	0.76312
MOTA	0.60940	0.74265	0.71591
FP	0.25110	0.02206	0.02110
TP	0.86050	0.76471	0.73701
FN	0.13950	0.23529	0.26299
ID switch	0	0	0

##### B.1.1.3 Nivel 3

Métrica	Secuencia 1	Secuencia 2	Secuencia 3	Secuencia 4	Secuencia 5
MOTP	0.70501	0.69448	0.70604	0.74787	0.70836
MOTA	0.73835	0.59339	0.80101	0.85714	0.38136
FP	0.03308	0.00000	0.00000	0.00000	0.13983
TP	0.77143	0.59339	0.80101	0.85714	0.52119
FN	0.22707	0.40467	0.19899	0.13946	0.47881
ID switch	1	2	0	1	0

##### B.1.1.4 Nivel 4

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.69214	0.71182	0.70263
MOTA	0.78922	0.54561	0.21088
FP	0.04412	0.20270	0.37415
TP	0.83333	0.74831	0.58503
FN	0.16667	0.24662	0.37415
ID switch	0	3	6

##### B.1.1.5 Nivel 5

Métrica	Secuencia 1	Secuencia 2
MOTP	0.67730	0.69452
MOTA	0.39024	0.18716
FP	0.20976	0.29541
TP	0.60000	0.48257
FN	0.40000	0.51376
ID switch	0	2

##### B.1.1.6 Nivel 6

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.71084	0.67750	0.69232
MOTA	0.40177	0.31597	0.24741
FP	0.16991	0.24370	0.15407
TP	0.57168	0.55966	0.40148
FN	0.40708	0.42353	0.59111
ID switch	12	10	5

## B.2. Segunda modificación

### B.2.1 Resultados obtenidos para cada nivel del dataset

#### B.2.1.1 Nivel 1

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.78435	0.72724	0.69564
MOTA	0.91304	0.66554	0.91994
FP	0.00242	0.27365	0.01570
TP	0.91546	0.93919	0.93564
FN	0.08454	0.06081	0.06436
ID switch	0	0	0

#### B.2.1.2 Nivel 2

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.77348	0.68711	0.76779
MOTA	0.61087	0.71324	0.69643
FP	0.25110	0.03677	0.03084
TP	0.86197	0.75	0.72727
FN	0.13803	0.25	0.27273
ID switch	0	0	0

#### B.2.1.3 Nivel 3

Métrica	Secuencia 1	Secuencia 2	Secuencia 3	Secuencia 4	Secuencia 5
MOTP	0.70511	0.69448	0.70604	0.74787	0.65716
MOTA	0.74135	0.59339	0.80101	0.85714	0.42373
FP	0.03158	0.00000	0.00000	0.00000	0.11864
TP	0.77293	0.59339	0.80101	0.85714	0.54237
FN	0.22556	0.40467	0.19899	0.13946	0.45763
ID switch	1	2	0	1	0

#### B.2.1.4 Nivel 4

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.69214	0.71182	0.68220
MOTA	0.78922	0.54561	0.27211
FP	0.04412	0.20270	0.34694
TP	0.83333	0.74831	0.61905
FN	0.16667	0.24662	0.32653
ID switch	0	3	8

#### B.2.1.5 Nivel 5

Métrica	Secuencia 1	Secuencia 2
MOTP	0.67324	0.68855
MOTA	0.48780	0.11560
FP	0.16098	0.32294
TP	0.64878	0.43853
FN	0.35122	0.55596
ID switch	0	3

#### B.2.1.6 Nivel 6

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.71018	0.68841	0.67111
MOTA	0.48673	0.41345	0.26963
FP	0.15575	0.19832	0.16148
TP	0.64248	0.61176	0.43111
FN	0.34336	0.37311	0.55852
ID switch	8	9	7

## B.3. Tercera modificación

### B.3.1 Resultados obtenidos para cada nivel del dataset

#### B.3.1.1 Nivel 1

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.78521	0.72936	0.69564
MOTA	0.92754	0.68581	0.91994
FP	0.00242	0.27365	0.01570
TP	0.92995	0.95946	0.93564
FN	0.07005	0.04054	0.06436
ID switch	0	0	0

#### B.3.1.2 Nivel 2

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.77325	0.68896	0.76846
MOTA	0.61233	0.75000	0.70617
FP	0.25110	0.03677	0.03084
TP	0.86344	0.78676	0.73701
FN	0.13656	0.21324	0.26299
ID switch	0	0	0

### B.3.1.3 Nivel 3

Métrica	Secuencia 1	Secuencia 2	Secuencia 3	Secuencia 4	Secuencia 5
MOTP	0.70637	0.69448	0.70604	0.74787	0.65326
MOTA	0.75188	0.59339	0.80101	0.85714	0.48305
FP	0.03158	0.00000	0.00000	0.00000	0.12288
TP	0.78346	0.59339	0.80101	0.85714	0.60593
FN	0.21504	0.40467	0.19899	0.13946	0.39407
ID switch	1	2	0	1	0

### B.3.1.4 Nivel 4

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.69214	0.71243	0.68210
MOTA	0.78922	0.56926	0.28571
FP	0.04412	0.20270	0.34694
TP	0.83333	0.77196	0.63265
FN	0.16667	0.22128	0.31293
ID switch	0	4	8

### B.3.1.5 Nivel 5

Métrica	Secuencia 1	Secuencia 2
MOTP	0.67324	0.68356
MOTA	0.45854	0.13028
FP	0.19024	0.34128
TP	0.64878	0.47156
FN	0.35122	0.52294
ID switch	0	3

### B.3.1.6 Nivel 6

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.70622	0.68823	0.67097
MOTA	0.52212	0.41176	0.28148
FP	0.16283	0.21681	0.17037
TP	0.68496	0.62857	0.45185
FN	0.30088	0.35462	0.53630
ID switch	8	10	8

## B.4. Cuarta modificación

### B.4.1 Resultados obtenidos para cada nivel del dataset en función del parámetro $\beta$

#### B.4.1.1 $\beta = 0.3$

##### B.4.1.1.1 Nivel 1

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.78521	0.72936	0.69564
MOTA	0.92754	0.68581	0.91994
FP	0.00242	0.27365	0.01570
TP	0.92995	0.95946	0.93564
FN	0.07005	0.04054	0.06436
ID switch	0	0	0

##### B.4.1.1.2 Nivel 2

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.77325	0.68896	0.76846
MOTA	0.61233	0.75000	0.70617
FP	0.25110	0.03677	0.03084
TP	0.86344	0.78676	0.73701
FN	0.13656	0.21324	0.26299
ID switch	0	0	0

##### B.4.1.1.3 Nivel 3

Métrica	Secuencia 1	Secuencia 2	Secuencia 3	Secuencia 4	Secuencia 5
MOTP	0.70637	0.69448	0.70604	0.74787	0.65326
MOTA	0.75188	0.59339	0.80101	0.85714	0.48305
FP	0.03158	0.00000	0.00000	0.00000	0.12288
TP	0.78346	0.59339	0.80101	0.85714	0.60593
FN	0.21504	0.40467	0.19899	0.13946	0.39407
ID switch	1	2	0	1	0

#### B.4.1.1.4 Nivel 4

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.69214	0.71243	0.68210
MOTA	0.78922	0.56926	0.28571
FP	0.04412	0.20270	0.34694
TP	0.83333	0.77196	0.63265
FN	0.16667	0.22128	0.31293
ID switch	0	4	8

#### B.4.1.1.5 Nivel 5

Métrica	Secuencia 1	Secuencia 2
MOTP	0.67324	0.68356
MOTA	0.45854	0.13028
FP	0.19024	0.34128
TP	0.64878	0.47156
FN	0.35122	0.52294
ID switch	0	3

#### B.4.1.1.6 Nivel 6

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.70622	0.68823	0.67097
MOTA	0.52212	0.41176	0.28148
FP	0.16283	0.21681	0.17037
TP	0.68496	0.62857	0.45185
FN	0.30088	0.35462	0.53630
ID switch	8	10	8

#### B.4.1.2 $\beta = 0.4$

##### B.4.1.2.1 Nivel 1

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.78668	0.72936	0.69707
MOTA	0.90338	0.68581	0.91680
FP	0.01449	0.27365	0.01727
TP	0.91787	0.95946	0.93407
FN	0.08213	0.04054	0.06593
ID switch	0	0	0

##### B.4.1.2.2 Nivel 2

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.77325	0.69153	0.76861
MOTA	0.61233	0.75000	0.69463
FP	0.25110	0.03677	0.03571
TP	0.86344	0.78676	0.73214
FN	0.13656	0.21324	0.26786
ID switch	0	0	0

##### B.4.1.2.3 Nivel 3

Métrica	Secuencia 1	Secuencia 2	Secuencia 3	Secuencia 4	Secuencia 5
MOTP	0.70827	0.69448	0.70604	0.74988	0.65577
MOTA	0.75188	0.59339	0.80101	0.84354	0.51695
FP	0.03158	0.00000	0.00000	0.00000	0.10593
TP	0.78346	0.59339	0.80101	0.84354	0.62288
FN	0.21504	0.40467	0.19899	0.15306	0.37712
ID switch	1	2	0	1	0

##### B.4.1.2.4 Nivel 4

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.69214	0.71507	0.68808
MOTA	0.78922	0.56419	0.32653
FP	0.04412	0.20270	0.31973
TP	0.83333	0.76689	0.64626
FN	0.16667	0.22635	0.31973
ID switch	0	4	5

### B.4.1.2.5 Nivel 5

Métrica	Secuencia 1	Secuencia 2
MOTP	0.67232	0.67451
MOTA	0.49756	0.23303
FP	0.17073	0.29358
TP	0.66829	0.52661
FN	0.33171	0.46972
ID switch	0	2

### B.4.1.2.6 Nivel 6

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.71711	0.69242	0.67309
MOTA	0.52743	0.45882	0.29333
FP	0.14513	0.18824	0.17630
TP	0.67257	0.64706	0.46963
FN	0.30796	0.32773	0.51407
ID switch	11	15	11

### B.4.1.3 $\beta = 0.5$

#### B.4.1.3.1 Nivel 1

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.78651	0.72936	0.69660
MOTA	0.90338	0.68581	0.91366
FP	0.01449	0.27365	0.01884
TP	0.91787	0.95946	0.93250
FN	0.08213	0.04054	0.06750
ID switch	0	0	0

#### B.4.1.3.2 Nivel 2

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.77325	0.69760	0.76810
MOTA	0.61233	0.76471	0.69643
FP	0.25110	0.02941	0.03571
TP	0.86344	0.79412	0.73214
FN	0.13656	0.20588	0.26786
ID switch	0	0	0

#### B.4.1.3.3 Nivel 3

Métrica	Secuencia 1	Secuencia 2	Secuencia 3	Secuencia 4	Secuencia 5
MOTP	0.70691	0.69129	0.70604	0.74988	0.66606
MOTA	0.74586	0.59144	0.80101	0.84354	0.55932
FP	0.03459	0.00097	0.00000	0.00000	0.00000
TP	0.78045	0.59241	0.80101	0.84354	0.64407
FN	0.21805	0.40564	0.19899	0.15306	0.35593
ID switch	1	2	0	1	0

#### B.4.1.3.4 Nivel 4

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.69214	0.71033	0.67960
MOTA	0.78922	0.57939	0.40816
FP	0.04412	0.16723	0.23810
TP	0.83333	0.74662	0.64626
FN	0.16667	0.25000	0.33333
ID switch	0	2	3

#### B.4.1.3.5 Nivel 5

Métrica	Secuencia 1	Secuencia 2
MOTP	0.67268	0.69396
MOTA	0.49268	0.28807
FP	0.16585	0.26606
TP	0.65854	0.55413
FN	0.34146	0.44037
ID switch	0	3

#### B.4.1.3.6 Nivel 6

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.71325	0.68971	0.66746
MOTA	0.55044	0.43529	0.31556
FP	0.14336	0.19160	0.15560
TP	0.69381	0.62689	0.47111
FN	0.2885	0.35294	0.51852
ID switch	10	12	7

### B.4.1.4 $\beta = 0.6$

#### B.4.1.4.1 Nivel 1

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.78544	0.72936	0.69564
MOTA	0.89372	0.68581	0.91994
FP	0.01932	0.27365	0.01570
TP	0.91304	0.95946	0.93564
FN	0.08696	0.04054	0.06436
ID switch	0	0	0

#### B.4.1.4.2 Nivel 2

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.77203	0.69279	0.76786
MOTA	0.61233	0.77941	0.70617
FP	0.25110	0.02206	0.03084
TP	0.86344	0.80147	0.73701
FN	0.13656	0.19853	0.26299
ID switch	0	0	0

#### B.4.1.4.3 Nivel 3

Métrica	Secuencia 1	Secuencia 2	Secuencia 3	Secuencia 4	Secuencia 5
MOTP	0.69824	0.68412	0.70507	0.74580	0.67374
MOTA	0.72180	0.58949	0.80101	0.84354	0.57627
FP	0.04662	0.00195	0.00000	0.00000	0.07627
TP	0.76842	0.59144	0.80101	0.84354	0.65254
FN	0.23008	0.40661	0.19899	0.15306	0.34746
ID switch	1	2	0	1	0

#### B.4.1.4.4 Nivel 4

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.69327	0.71378	0.65637
MOTA	0.79412	0.60980	0.36054
FP	0.04167	0.15203	0.23129
TP	0.83578	0.76182	0.59184
FN	0.16422	0.23480	0.38776
ID switch	0	2	3

#### B.4.1.4.5 Nivel 5

Métrica	Secuencia 1	Secuencia 4
MOTP	0.66176	0.69075
MOTA	0.62927	0.35413
FP	0.10244	0.23303
TP	0.73171	0.58716
FN	0.26829	0.40734
ID switch	0	3

#### B.4.1.4.6 Nivel 6

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.71778	0.69875	0.67727
MOTA	0.62655	0.45210	0.31852
FP	0.10088	0.18319	0.15111
TP	0.72743	0.63529	0.46963
FN	0.25310	0.34454	0.52296
ID switch	11	12	5

### B.4.1.5 $\beta = 0.7$

#### B.4.1.5.1 Nivel 1

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.78341	0.72794	0.69730
MOTA	0.90338	0.68581	0.91994
FP	0.01449	0.27365	0.01570
TP	0.91787	0.95946	0.93564
FN	0.08213	0.04054	0.06436
ID switch	0	0	0

#### B.4.1.5.2 Nivel 2

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.76940	0.70430	0.76532
MOTA	0.61233	0.76471	0.70942
FP	0.25110	0.02941	0.02922
TP	0.86344	0.79412	0.73864
FN	0.13656	0.20588	0.26136
ID switch	0	0	0

### B.4.1.5.3 Nivel 3

Métrica	Secuencia 1	Secuencia 2	Secuencia 3	Secuencia 4	Secuencia 5
MOTP	0.69896	0.68213	0.70876	0.75006	0.67891
MOTA	0.72782	0.59144	0.80101	0.84354	0.60169
FP	0.04361	0.00097	0.00000	0.00000	0.06356
TP	0.77143	0.59241	0.80101	0.84354	0.66525
FN	0.22707	0.40564	0.19899	0.15306	0.33475
ID switch	1	2	0	1	0

### B.4.1.5.4 Nivel 4

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.76961	0.71977	0.67072
MOTA	0.69280	0.58953	0.34694
FP	0.05392	0.16216	0.23810
TP	0.82353	0.75169	0.58503
FN	0.17647	0.24493	0.39456
ID switch	0	2	3

### B.4.1.5.5 Nivel 5

Métrica	Secuencia 1	Secuencia 2
MOTP	0.66988	0.69594
MOTA	0.64878	0.36330
FP	0.09268	0.23119
TP	0.74146	0.59450
FN	0.25854	0.40183
ID switch	0	2

### B.4.1.5.6 Nivel 6

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.71862	0.69197	0.68009
MOTA	0.61947	0.45546	0.34963
FP	0.09381	0.18319	0.13481
TP	0.71327	0.63866	0.48444
FN	0.27257	0.34118	0.50667
ID switch	8	12	6

## B.4.2 Comparativa de resultados medios obtenidos para cada nivel del dataset en función del parámetro $\beta$

Métrica	Nivel 1					Nivel 2				
	$\beta=0.3$	$\beta=0.4$	$\beta=0.5$	$\beta=0.6$	$\beta=0.7$	$\beta=0.3$	$\beta=0.4$	$\beta=0.5$	$\beta=0.6$	$\beta=0.7$
MOTP	0.73674	0.73770	0.73749	0.73681	0.73622	0.74356	0.74446	0.74632	0.74423	0.74634
MOTA	0.84443	0.83533	0.83428	0.83316	0.83638	0.68950	0.68565	0.69116	0.69930	0.69549
FP	0.09726	0.10180	0.10233	0.10289	0.10128	0.10624	0.10786	0.12108	0.10133	0.10324
TP	0.94168	0.93713	0.93661	0.93605	0.93766	0.79574	0.79411	0.77295	0.80064	0.79873
FN	0.05832	0.06287	0.06339	0.06395	0.06234	0.20426	0.20589	0.222705	0.19936	0.20127
ID switch	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Métrica	Nivel 3					Nivel 4				
	$\beta=0.3$	$\beta=0.4$	$\beta=0.5$	$\beta=0.6$	$\beta=0.7$	$\beta=0.3$	$\beta=0.4$	$\beta=0.5$	$\beta=0.6$	$\beta=0.7$
MOTP	0.70160	0.70289	0.70404	0.70140	0.70376	0.69556	0.69843	0.69402	0.68781	0.72003
MOTA	0.69729	0.70135	0.70823	0.70642	0.71310	0.54806	0.55998	0.59226	0.58815	0.54309
FP	0.03089	0.02750	0.00711	0.02497	0.02163	0.19792	0.18885	0.14982	0.14166	0.15140
TP	0.72819	0.72886	0.73230	0.73139	0.61439	0.74598	0.74883	0.74207	0.72981	0.72008
FN	0.27045	0.26978	0.26633	0.26724	0.33002	0.23363	0.23758	0.25000	0.26226	0.27199
ID switch	0.80	0.80	1.00	0.80	0.80	4.00	3.00	0.80	0.80	0.80

Métrica	Nivel 5					Nivel 6				
	$\beta=0.3$	$\beta=0.4$	$\beta=0.5$	$\beta=0.6$	$\beta=0.7$	$\beta=0.3$	$\beta=0.4$	$\beta=0.5$	$\beta=0.6$	$\beta=0.7$
MOTP	0.67840	0.67342	0.68332	0.67626	0.68291	0.68847	0.69421	0.69014	0.69793	0.69689
MOTA	0.29441	0.36530	0.39038	0.49170	0.50604	0.40512	0.42653	0.43376	0.46572	0.47485
FP	0.26576	0.23216	0.21560	0.16774	0.16194	0.18334	0.16989	0.16352	0.14506	0.13727
TP	0.56017	0.59745	0.60634	0.65944	0.66798	0.58846	0.59642	0.59727	0.61078	0.61212
FN	0.43708	0.40072	0.39092	0.33782	0.33019	0.39727	0.38325	0.38665	0.37353	0.37347
ID switch	1.50	1.00	1.50	1.50	1.00	8.67	10.00	9.67	9.33	8.67

## B.5. Quinta modificación

### B.5.1 Resultados obtenidos para cada nivel del dataset

#### B.5.1.1.1 Nivel 1

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.78544	0.72936	0.69564
MOTA	0.89372	0.68581	0.91994
FP	0.01932	0.27365	0.01570
TP	0.91304	0.95946	0.93564
FN	0.08696	0.04054	0.06436
ID switch	0	0	0

#### B.5.1.1.2 Nivel 2

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.77203	0.69279	0.76834
MOTA	0.61233	0.77941	0.70942
FP	0.25110	0.02206	0.02110
TP	0.86344	0.80147	0.73052
FN	0.13656	0.19853	0.26948
ID switch	0	0	0

#### B.5.1.1.3 Nivel 3

Métrica	Secuencia 1	Secuencia 2	Secuencia 3	Secuencia 4	Secuencia 5
MOTP	0.69824	0.68412	0.70507	0.74580	0.67374
MOTA	0.72180	0.58949	0.80101	0.84354	0.57627
FP	0.04662	0.00195	0.00000	0.00000	0.07627
TP	0.76842	0.59144	0.80101	0.84354	0.65254
FN	0.23008	0.40661	0.19899	0.15306	0.34746
ID switch	1	2	0	1	0

#### B.5.1.1.4 Nivel 4

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.69327	0.71378	0.65637
MOTA	0.81863	0.62838	0.36054
FP	0.01716	0.13345	0.23129
TP	0.83578	0.76182	0.59184
FN	0.16422	0.23480	0.38776
ID switch	0	2	3

#### B.5.1.1.5 Nivel 5

Métrica	Secuencia 1	Secuencia 2
MOTP	0.66218	0.69117
MOTA	0.63902	0.39450
FP	0.08781	0.18899
TP	0.72683	0.58349
FN	0.27317	0.41101
ID switch	0	3

#### B.5.1.1.6 Nivel 6

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.71778	0.69875	0.67727
MOTA	0.63363	0.45882	0.33037
FP	0.09381	0.17647	0.13926
TP	0.72743	0.63529	0.46963
FN	0.25310	0.34454	0.52296
ID switch	11	12	5

## B.6. Sexta modificación

### B.6.1 Resultados obtenidos para cada nivel del dataset en función del parámetro $\alpha$

#### B.6.1.1 $\alpha = 0.0$

##### B.6.1.1.1 Nivel 1

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.76552	0.72262	0.68821
MOTA	0.89372	0.67905	0.89168
FP	0.01932	0.27703	0.02983
TP	0.91304	0.95608	0.92151
FN	0.08696	0.04392	0.07849
ID switch	0	0	0

##### B.6.1.1.2 Nivel 2

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.76093	0.72341	0.75624
MOTA	0.61233	0.69118	0.70617
FP	0.25110	0.09559	0.02273
TP	0.86344	0.78676	0.72890
FN	0.13656	0.20588	0.27110
ID switch	0	1	0

##### B.6.1.1.3 Nivel 3

Métrica	Secuencia 1	Secuencia 2	Secuencia 3	Secuencia 4	Secuencia 5
MOTP	0.69634	0.68693	0.69730	0.74018	0.70876
MOTA	0.70977	0.59339	0.79093	0.58844	0.61017
FP	0.05263	0.00000	0.00504	0.24490	0.06356
TP	0.76241	0.59339	0.79597	0.83333	0.67373
FN	0.23609	0.40467	0.20403	0.14626	0.32203
ID switch	1	2	0	6	1

##### B.6.1.1.4 Nivel 4

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.68421	0.72598	0.69043
MOTA	0.79902	0.60811	0.33333
FP	0.02700	0.14358	0.24490
TP	0.82598	0.75169	0.57823
FN	0.17402	0.24493	0.40136
ID switch	0	2	3

##### B.6.1.1.5 Nivel 5

Métrica	Secuencia 1	Secuencia 2
MOTP	0.70223	0.70981
MOTA	0.64390	0.50642
FP	0.08293	0.13211
TP	0.72683	0.63853
FN	0.27317	0.35596
ID switch	0	3

##### B.6.1.1.6 Nivel 6

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.70694	0.68402	0.71585
MOTA	0.60177	0.54454	0.45926
FP	0.11150	0.13782	0.07259
TP	0.71327	0.68235	0.53185
FN	0.27080	0.29748	0.45630
ID switch	9	12	8

#### B.6.1.2 $\alpha = 0.1$

##### B.6.1.2.1 Nivel 1

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.77139	0.72507	0.69047
MOTA	0.89372	0.67905	0.90424
FP	0.01932	0.27703	0.02355
TP	0.91304	0.95608	0.92779
FN	0.08696	0.04392	0.07221
ID switch	0	0	0

##### B.6.1.2.2 Nivel 2

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.76469	0.72451	0.75923
MOTA	0.61233	0.70588	0.70617
FP	0.25110	0.08824	0.02273
TP	0.86344	0.79412	0.72890
FN	0.13656	0.19853	0.27110
ID switch	0	1	0

### B.6.1.2.3 Nivel 3

Métrica	Secuencia 1	Secuencia 2	Secuencia 3	Secuencia 4	Secuencia 5
MOTP	0.69973	0.68798	0.70020	0.74090	0.70714
MOTA	0.71880	0.59339	0.80101	0.84354	0.63559
FP	0.04812	0.00000	0.00000	0.01361	0.04661
TP	0.76692	0.59339	0.80101	0.85714	0.68220
FN	0.23158	0.40467	0.19899	0.13946	0.31780
ID switch	1	2	0	1	0

### B.6.1.2.4 Nivel 4

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.68656	0.72572	0.68835
MOTA	0.80882	0.63176	0.34694
FP	0.02206	0.13176	0.23810
TP	0.83088	0.76351	0.58503
FN	0.16912	0.23311	0.39456
ID switch	0	2	3

### B.6.1.2.5 Nivel 5

Métrica	Secuencia 1	Secuencia 2
MOTP	0.70242	0.70984
MOTA	0.64390	0.51376
FP	0.08293	0.12844
TP	0.72683	0.64220
FN	0.27317	0.35229
ID switch	0	3

### B.6.1.2.6 Nivel 6

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.71429	0.69039	0.71192
MOTA	0.62655	0.55462	0.45926
FP	0.09912	0.13277	0.07259
TP	0.72566	0.68739	0.53185
FN	0.25841	0.29244	0.45630
ID switch	9	12	8

## B.6.1.3 $\alpha = 0.2$

### B.6.1.3.1 Nivel 1

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.77681	0.72713	0.69276
MOTA	0.89372	0.67905	0.91052
FP	0.01932	0.27703	0.02041
TP	0.91304	0.95608	0.93093
FN	0.08696	0.04392	0.06907
ID switch	0	0	0

### B.6.1.3.2 Nivel 2

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.76801	0.72170	0.76283
MOTA	0.61233	0.80882	0.70292
FP	0.25110	0.00735	0.02435
TP	0.86344	0.81618	0.72727
FN	0.13656	0.18382	0.27273
ID switch	0	0	0

### B.6.1.3.3 Nivel 3

Métrica	Secuencia 1	Secuencia 2	Secuencia 3	Secuencia 4	Secuencia 5
MOTP	0.70110	0.68817	0.70339	0.74368	0.70082
MOTA	0.73083	0.59339	0.80101	0.86395	0.63559
FP	0.04211	0.00000	0.00000	0.00340	0.04661
TP	0.77293	0.59339	0.80101	0.86735	0.68220
FN	0.22556	0.40467	0.19899	0.12925	0.31780
ID switch	1	2	0	1	0

#### B.6.1.3.4 Nivel 4

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.68894	0.72625	0.68226
MOTA	0.81373	0.63514	0.34694
FP	0.01961	0.13007	0.23810
TP	0.83333	0.76520	0.58503
FN	0.16667	0.23142	0.39456
ID switch	0	2	3

#### B.6.1.3.5 Nivel 5

Métrica	Secuencia 1	Secuencia 2
MOTP	0.69821	0.70802
MOTA	0.64390	0.50092
FP	0.08230	0.13578
TP	0.72683	0.63670
FN	0.27317	0.35780
ID switch	0	3

#### B.6.1.3.6 Nivel 6

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.72050	0.69641	0.70418
MOTA	0.64425	0.54454	0.44444
FP	0.09027	0.13782	0.08148
TP	0.73451	0.68235	0.52593
FN	0.24956	0.29748	0.46519
ID switch	9	12	6

### B.6.1.4 $\alpha = 0.3$

#### B.6.1.4.1 Nivel 1

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.78098	0.72779	0.69459
MOTA	0.89372	0.68581	0.91366
FP	0.01932	0.27365	0.01884
TP	0.91304	0.95946	0.93250
FN	0.08696	0.04054	0.06750
ID switch	0	0	0

#### B.6.1.4.2 Nivel 2

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.77037	0.71623	0.76397
MOTA	0.61233	0.80882	0.70617
FP	0.25110	0.00735	0.02273
TP	0.86344	0.81618	0.72890
FN	0.13656	0.18382	0.27110
ID switch	0	0	0

#### B.6.1.4.3 Nivel 3

Métrica	Secuencia 1	Secuencia 2	Secuencia 3	Secuencia 4	Secuencia 5
MOTP	0.70286	0.68769	0.70519	0.74641	0.61864
MOTA	0.72481	0.59144	0.80101	0.86395	0.69383
FP	0.04511	0.00097	0.00000	0.00340	0.05509
TP	0.76992	0.59241	0.80101	0.86735	0.67373
FN	0.22857	0.40564	0.19899	0.12925	0.32627
ID switch	1	2	0	1	0

#### B.6.1.4.4 Nivel 4

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.69076	0.72420	0.67371
MOTA	0.81863	0.63514	0.36054
FP	0.01716	0.13007	0.23129
TP	0.83578	0.76520	0.59184
FN	0.16422	0.23142	0.38776
ID switch	0	2	3

### B.6.1.4.5 Nivel 5

Métrica	Secuencia 1	Secuencia 2
MOTP	0.68697	0.70238
MOTA	0.65366	0.48257
FP	0.07805	0.14495
TP	0.73171	0.62752
FN	0.26829	0.36697
ID switch	0	3

### B.6.1.4.6 Nivel 6

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.72322	0.69671	0.69467
MOTA	0.64779	0.54790	0.41630
FP	0.08673	0.13613	0.09482
TP	0.73451	0.68403	0.51111
FN	0.24602	0.29580	0.47852
ID switch	11	12	7

### B.6.1.5 $\alpha = 0.4$

#### B.6.1.5.1 Nivel 1

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.78395	0.72883	0.69567
MOTA	0.89372	0.68581	0.91680
FP	0.01932	0.27365	0.01727
TP	0.91304	0.95946	0.93407
FN	0.05696	0.04054	0.06593
ID switch	0	0	0

#### B.6.1.5.2 Nivel 2

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.77179	0.70507	0.76255
MOTA	0.61233	0.80882	0.71266
FP	0.25110	0.00735	0.01948
TP	0.86344	0.81618	0.73214
FN	0.13656	0.18382	0.26786
ID switch	0	0	0

#### B.6.1.5.3 Nivel 3

Métrica	Secuencia 1	Secuencia 2	Secuencia 3	Secuencia 4	Secuencia 5
MOTP	0.70225	0.68631	0.70571	0.74665	0.68363
MOTA	0.72180	0.58949	0.80101	0.87075	0.60169
FP	0.04662	0.00195	0.00000	0.00000	0.06356
TP	0.76842	0.59144	0.80101	0.87075	0.66525
FN	0.23008	0.40661	0.19899	0.12585	0.33475
ID switch	1	2	0	1	0

#### B.6.1.5.4 Nivel 4

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.69261	0.72064	0.66458
MOTA	0.81863	0.62838	0.37415
FP	0.01716	0.13345	0.22449
TP	0.83578	0.76182	0.59864
FN	0.16422	0.23480	0.38095
ID switch	0	2	3

#### B.6.1.5.5 Nivel 5

Métrica	Secuencia 1	Secuencia 2
MOTP	0.67665	0.69875
MOTA	0.62439	0.43119
FP	0.09268	0.17064
TP	0.71707	0.60183
FN	0.28293	0.39266
ID switch	0	3

#### B.6.1.5.6 Nivel 6

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.72150	0.69997	0.68369
MOTA	0.64779	0.50420	0.38667
FP	0.08673	0.15798	0.10963
TP	0.73451	0.66218	0.49630
FN	0.24602	0.31765	0.49333
ID switch	11	12	7

### B.6.1.6 $\alpha = 0.5$

#### B.6.1.6.1 Nivel 1

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.78521	0.72936	0.69564
MOTA	0.92754	0.68581	0.91994
FP	0.00242	0.27365	0.01570
TP	0.92995	0.95946	0.93564
FN	0.07005	0.04054	0.06436
ID switch	0	0	0

#### B.6.1.6.2 Nivel 2

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.77203	0.69279	0.76834
MOTA	0.61233	0.77941	0.70942
FP	0.25110	0.02206	0.02110
TP	0.86344	0.80147	0.73052
FN	0.13656	0.19853	0.26948
ID switch	0	0	0

#### B.6.1.6.3 Nivel 3

Métrica	Secuencia 1	Secuencia 2	Secuencia 3	Secuencia 4	Secuencia 5
MOTP	0.69824	0.68412	0.70507	0.74580	0.67374
MOTA	0.72180	0.58949	0.80101	0.84354	0.57627
FP	0.04662	0.00195	0.00000	0.00000	0.07627
TP	0.76842	0.59144	0.80101	0.84354	0.65254
FN	0.23008	0.40661	0.19899	0.15306	0.34746
ID switch	1	2	0	1	0

#### B.6.1.6.4 Nivel 4

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.69327	0.71378	0.65637
MOTA	0.81863	0.62838	0.36054
FP	0.01716	0.13345	0.23129
TP	0.83578	0.76182	0.59184
FN	0.16422	0.23480	0.38776
ID switch	0	2	3

#### B.6.1.6.5 Nivel 5

Métrica	Secuencia 1	Secuencia 2
MOTP	0.66218	0.69117
MOTA	0.63902	0.39450
FP	0.08781	0.18899
TP	0.72683	0.58349
FN	0.27317	0.41101
ID switch	0	3

#### B.6.1.6.6 Nivel 6

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.71778	0.69875	0.67727
MOTA	0.63363	0.45882	0.33037
FP	0.09381	0.17647	0.13926
TP	0.72743	0.63529	0.46963
FN	0.25310	0.34454	0.52296
ID switch	11	12	5

### B.6.1.7 $\alpha = 0.6$

#### B.6.1.7.1 Nivel 1

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.78539	0.72908	0.69488
MOTA	0.89372	0.68581	0.91680
FP	0.01932	0.27365	0.01727
TP	0.91304	0.95946	0.93407
FN	0.08696	0.04054	0.06593
ID switch	0	0	0

#### B.6.1.7.2 Nivel 2

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.77077	0.67702	0.75739
MOTA	0.61233	0.76471	0.71266
FP	0.25110	0.02941	0.01948
TP	0.86344	0.79412	0.73214
FN	0.13656	0.20588	0.26786
ID switch	0	0	0

### B.6.1.7.3 Nivel 3

Métrica	Secuencia 1	Secuencia 2	Secuencia 3	Secuencia 4	Secuencia 5
MOTP	0.69565	0.68242	0.70305	0.74848	0.66643
MOTA	0.71880	0.58366	0.80101	0.84694	0.52542
FP	0.04812	0.00486	0.00000	0.00000	0.10169
TP	0.76692	0.58852	0.80101	0.84694	0.62712
FN	0.23158	0.40953	0.19899	0.14966	0.37288
ID switch	1	2	0	1	0

### B.6.1.7.4 Nivel 4

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.69355	0.70679	0.64694
MOTA	0.81373	0.61824	0.33333
FP	0.01961	0.13851	0.24490
TP	0.83333	0.75676	0.57823
FN	0.16667	0.23986	0.40136
ID switch	0	2	3

### B.6.1.7.5 Nivel 5

Métrica	Secuencia 1	Secuencia 2
MOTP	0.65224	0.68645
MOTA	0.59024	0.32110
FP	0.11220	0.22569
TP	0.70244	0.54679
FN	0.29756	0.44771
ID switch	0	3

### B.6.1.7.6 Nivel 6

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.71173	0.69320	0.67245
MOTA	0.59823	0.43866	0.26074
FP	0.11327	0.18655	0.17333
TP	0.71150	0.62521	0.43407
FN	0.26726	0.35462	0.55704
ID switch	12	12	6

### B.6.1.8 $\alpha = 0.7$

#### B.6.1.8.1 Nivel 1

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.78368	0.72835	0.69299
MOTA	0.89855	0.68581	0.91680
FP	0.01691	0.27365	0.01727
TP	0.91546	0.95946	0.93407
FN	0.08454	0.04054	0.06593
ID switch	0	0	0

#### B.6.1.8.2 Nivel 2

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.76632	0.66761	0.75310
MOTA	0.57856	0.69118	0.71266
FP	0.27900	0.06618	0.01948
TP	0.85756	0.75735	0.73214
FN	0.13950	0.24265	0.26786
ID switch	2	0	0

#### B.6.1.8.3 Nivel 3

Métrica	Secuencia 1	Secuencia 2	Secuencia 3	Secuencia 4	Secuencia 5
MOTP	0.68992	0.68007	0.70014	0.74640	0.66424
MOTA	0.71880	0.57977	0.80101	0.84694	0.44915
FP	0.04812	0.00681	0.00000	0.00000	0.13983
TP	0.76692	0.58658	0.80101	0.84694	0.58898
FN	0.23158	0.41148	0.19899	0.14966	0.41102
ID switch	1	2	0	1	0

#### B.6.1.8.4 Nivel 4

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.69168	0.69780	0.64613
MOTA	0.81863	0.61486	0.23810
FP	0.01716	0.14020	0.29252
TP	0.83578	0.75507	0.53061
FN	0.16422	0.24155	0.44898
ID switch	0	2	3

#### B.6.1.8.5 Nivel 5

Métrica	Secuencia 1	Secuencia 2
MOTP	0.64718	0.68458
MOTA	0.50244	0.23303
FP	0.15610	0.27156
TP	0.65854	0.50459
FN	0.34146	0.48807
ID switch	0	4

#### B.6.1.8.6 Nivel 6

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.70503	0.68683	0.66929
MOTA	0.54867	0.40168	0.20296
FP	0.13805	0.20504	0.19556
TP	0.68673	0.60672	0.39852
FN	0.29204	0.37311	0.58963
ID switch	12	12	8

#### B.6.1.9 $\alpha = 0.8$

##### B.6.1.9.1 Nivel 1

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.78151	0.72795	0.69089
MOTA	0.89855	0.61486	0.91052
FP	0.01691	0.34122	0.02041
TP	0.91546	0.95608	0.93093
FN	0.08454	0.04054	0.06907
ID switch	0	1	0

##### B.6.1.9.2 Nivel 2

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.76301	0.65680	0.74938
MOTA	0.57856	0.63235	0.70292
FP	0.27900	0.09559	0.02435
TP	0.85756	0.72794	0.72727
FN	0.13950	0.27206	0.27273
ID switch	2	0	0

##### B.6.1.9.3 Nivel 3

Métrica	Secuencia 1	Secuencia 2	Secuencia 3	Secuencia 4	Secuencia 5
MOTP	0.68407	0.67842	0.69650	0.73983	0.65698
MOTA	0.71278	0.57004	0.80101	0.86395	0.39831
FP	0.05113	0.01167	0.00000	0.00340	0.16525
TP	0.76391	0.58171	0.80101	0.86735	0.56356
FN	0.23459	0.41634	0.19899	0.12925	0.43644
ID switch	1	2	0	1	0

##### B.6.1.9.4 Nivel 4

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.68951	0.68953	0.64235
MOTA	0.80392	0.59797	0.18367
FP	0.01716	0.14865	0.31973
TP	0.82108	0.74662	0.50340
FN	0.17892	0.25000	0.47619
ID switch	0	2	3

### B.6.1.9.5 Nivel 5

Métrica	Secuencia 1	Secuencia 2
MOTP	0.64190	0.68327
MOTA	0.43415	0.15229
FP	0.19024	0.31193
TP	0.62439	0.46422
FN	0.37561	0.31193
ID switch	0	4

### B.6.1.9.6 Nivel 6

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.70052	0.67785	0.67280
MOTA	0.47965	0.35126	0.11704
FP	0.17345	0.23025	0.23852
TP	0.65310	0.58151	0.35556
FN	0.33097	0.39832	0.63259
ID switch	9	12	8

## B.6.2 Comparativa de resultados medios obtenidos para cada nivel del dataset en función del parámetro $\alpha$

Métrica	Nivel 1								
	$\alpha=0.0$	$\alpha=0.1$	$\alpha=0.2$	$\alpha=0.3$	$\alpha=0.4$	$\alpha=0.5$	$\alpha=0.6$	$\alpha=0.7$	$\alpha=0.8$
MOTP	0.73545	0.72898	0.73223	0.73445	0.73615	0.73681	0.73645	0.73501	0.73345
MOTA	0.82148	0.82567	0.82776	0.83106	0.83211	0.83316	0.83211	0.83372	0.80798
FP	0.10873	0.10663	0.10559	0.10394	0.10341	0.10289	0.10341	0.10261	0.12618
TP	0.93021	0.93230	0.93335	0.93500	0.93552	0.93605	0.93552	0.93633	0.93416
FN	0.06979	0.06770	0.06665	0.06500	0.05447	0.06400	0.06448	0.06367	0.06472
ID switch	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.33

Métrica	Nivel 2								
	$\alpha=0.0$	$\alpha=0.1$	$\alpha=0.2$	$\alpha=0.3$	$\alpha=0.4$	$\alpha=0.5$	$\alpha=0.6$	$\alpha=0.7$	$\alpha=0.8$
MOTP	0.74686	0.74948	0.75085	0.75019	0.74647	0.74439	0.73506	0.72901	0.72306
MOTA	0.66989	0.67479	0.70802	0.70911	0.71127	0.70039	0.69657	0.66080	0.63794
FP	0.12314	0.12069	0.09427	0.09373	0.09264	0.09809	0.09999	0.12155	0.13298
TP	0.79303	0.79549	0.80230	0.80284	0.80392	0.79848	0.79657	0.78235	0.77092
FN	0.20451	0.20206	0.19770	0.19716	0.19608	0.20152	0.20343	0.21667	0.22900
ID switch	0.33	0.33	0.00	0.00	0.00	0.00	0.00	0.67	0.67

Métrica	Nivel 3								
	$\alpha=0.0$	$\alpha=0.1$	$\alpha=0.2$	$\alpha=0.3$	$\alpha=0.4$	$\alpha=0.5$	$\alpha=0.6$	$\alpha=0.7$	$\alpha=0.8$
MOTP	0.70590	0.70719	0.70743	0.69216	0.70491	0.70139	0.69921	0.69615	0.69116
MOTA	0.65854	0.71847	0.72495	0.73501	0.71695	0.70642	0.69517	0.67913	0.66922
FP	0.07323	0.02167	0.01842	0.02091	0.02243	0.02497	0.03093	0.03895	0.04629
TP	0.73177	0.74013	0.74338	0.74088	0.73937	0.73139	0.72610	0.71809	0.71551
FN	0.26262	0.25850	0.25525	0.25774	0.25926	0.26724	0.27253	0.28055	0.28312
ID switch	2.00	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80

Métrica	Nivel 4								
	$\alpha=0.0$	$\alpha=0.1$	$\alpha=0.2$	$\alpha=0.3$	$\alpha=0.4$	$\alpha=0.5$	$\alpha=0.6$	$\alpha=0.7$	$\alpha=0.8$
MOTP	0.70021	0.70021	0.69915	0.69622	0.69261	0.68781	0.68243	0.67854	0.67380
MOTA	0.58015	0.59584	0.59860	0.60477	0.60705	0.60252	0.58843	0.55720	0.52852
FP	0.13849	0.13064	0.12926	0.12617	0.12503	0.12730	0.13434	0.14996	0.16185
TP	0.71863	0.72647	0.72785	0.73094	0.73208	0.72981	0.72277	0.70715	0.69037
FN	0.27344	0.26560	0.26422	0.26113	0.25999	0.26226	0.26930	0.28492	0.30170
ID switch	1.67	1.67	1.67	1.67	1.67	1.67	1.67	1.67	1.67

Métrica	Nivel 5								
	$\alpha=0.0$	$\alpha=0.1$	$\alpha=0.2$	$\alpha=0.3$	$\alpha=0.4$	$\alpha=0.5$	$\alpha=0.6$	$\alpha=0.7$	$\alpha=0.8$
MOTP	0.70602	0.70613	0.70312	0.69468	0.68770	0.67668	0.66935	0.66588	0.66259
MOTA	0.57516	0.57883	0.57241	0.56812	0.52779	0.51676	0.45567	0.36774	0.29322
FP	0.10752	0.10569	0.10904	0.11150	0.13166	0.13840	0.16895	0.21383	0.25109
TP	0.68268	0.68452	0.68177	0.67962	0.65945	0.65516	0.62462	0.58157	0.54431
FN	0.31457	0.31273	0.31549	0.31763	0.33780	0.34209	0.37264	0.41477	0.34377
ID switch	1.50	1.50	1.50	1.50	1.50	1.50	1.50	2.00	2.00

Métrica	Nivel 6								
	$\alpha=0.0$	$\alpha=0.1$	$\alpha=0.2$	$\alpha=0.3$	$\alpha=0.4$	$\alpha=0.5$	$\alpha=0.6$	$\alpha=0.7$	$\alpha=0.8$
MOTP	0.70227	0.70553	0.70703	0.70487	0.70172	0.69793	0.69246	0.68705	0.68372
MOTA	0.53519	0.54681	0.54441	0.53733	0.51289	0.47427	0.43254	0.38444	0.31598
FP	0.10730	0.10149	0.10319	0.10589	0.11811	0.13651	0.15772	0.17955	0.21407
TP	0.64249	0.64830	0.64760	0.64322	0.63000	0.61078	0.59026	0.56399	0.53006
FN	0.34153	0.33572	0.33741	0.34011	0.35233	0.37353	0.39297	0.41826	0.45396
ID switch	9.67	9.67	9.00	10.00	10.00	9.33	10.00	10.67	9.67



## C. Resultados finales obtenidos para la combinación del detector Latent SVM y el Filtro de Kalman

### C.1. Nivel 1

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.77779	0.72497	0.70108
MOTA	0.87923	0.65878	0.86028
FP	0.01932	0.27703	0.01884
TP	0.89855	0.93581	0.87912
FN	0.10145	0.06419	0.12088
ID switch	0	0	0

### C.2. Nivel 2

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.77014	0.71747	0.75927
MOTA	0.61380	0.72794	0.71916
FP	0.24816	0.02941	0.01136
TP	0.86197	0.75735	0.73052
FN	0.13803	0.24265	0.26948
ID switch	0	0	0

### C.3. Nivel 3

Métrica	Secuencia 1	Secuencia 2	Secuencia 3	Secuencia 4	Secuencia 5
MOTP	0.70193	0.68819	0.70563	0.74528	0.70391
MOTA	0.73985	0.58949	0.81108	0.86395	0.55932
FP	0.03308	0.00000	0.00000	0.00340	0.06356
TP	0.77293	0.58949	0.81108	0.86735	0.62288
FN	0.22556	0.40856	0.18892	0.12925	0.37712
ID switch	1	2	0	1	0

### C.4. Nivel 4

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.69246	0.73374	0.71728
MOTA	0.81863	0.63345	0.59864
FP	0.01471	0.12162	0.13605
TP	0.83333	0.75507	0.73469
FN	0.16667	0.24155	0.23810
ID switch	0	2	4

### C.5. Nivel 5

Métrica	Secuencia 1	Secuencia 2
MOTP	0.70718	0.71774
MOTA	0.58537	0.48624
FP	0.10244	0.18165
TP	0.68780	0.66789
FN	0.30732	0.33211
ID switch	1	0

### C.6. Nivel 6

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.68295	0.66124	0.69425
MOTA	0.67965	0.63193	0.50222
FP	0.05310	0.09076	0.02074
TP	0.73274	0.72269	0.52296
FN	0.25664	0.26723	0.46519
ID switch	6	6	8



## D. Refinamiento de los parámetros del Filtro de Partículas basado en color

### D.1. Asignación de $\sigma_x$ y $\sigma_y$ a un valor de 5 para las secuencias que pertenecen al *dataset* público PETS09

#### D.1.1 Resultados obtenidos para cada una de las secuencias de los niveles del *dataset*

##### D.1.1.1 Nivel 1

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.77681	0.72713	0.69276
MOTA	0.89372	0.67905	0.91052
FP	0.01932	0.27703	0.02041
TP	0.91304	0.95608	0.93093
FN	0.08696	0.04392	0.06907
ID switch	0	0	0

##### D.1.1.2 Nivel 2

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.76801	0.72589	0.76283
MOTA	0.61233	0.80882	0.70292
FP	0.25110	0.00735	0.02435
TP	0.86344	0.81618	0.72727
FN	0.13656	0.18382	0.27273
ID switch	0	0	0

##### D.1.1.3 Nivel 3

Métrica	Secuencia 1	Secuencia 2	Secuencia 3	Secuencia 4	Secuencia 5
MOTP	0.70110	0.68817	0.70339	0.74368	0.70360
MOTA	0.73083	0.59339	0.80101	0.86395	0.64407
FP	0.04211	0.00000	0.00000	0.00340	0.04237
TP	0.77293	0.59339	0.80101	0.86735	0.68644
FN	0.22556	0.40467	0.19899	0.12925	0.31356
ID switch	1	2	0	1	0

##### D.1.1.4 Nivel 4

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.68894	0.72625	0.68226
MOTA	0.81373	0.63514	0.34694
FP	0.01961	0.13007	0.23810
TP	0.83333	0.76520	0.58503
FN	0.16667	0.23142	0.39456
ID switch	0	2	3

##### D.1.1.5 Nivel 5

Métrica	Secuencia 1	Secuencia 2
MOTP	0.69854	0.71322
MOTA	0.71220	0.50092
FP	0.05366	0.13945
TP	0.76585	0.64037
FN	0.23415	0.35596
ID switch	0	2

##### D.1.1.6 Nivel 6

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.71718	0.69470	0.71308
MOTA	0.66726	0.55630	0.46815
FP	0.07611	0.13445	0.06667
TP	0.74336	0.69076	0.53481
FN	0.24425	0.29076	0.45333
ID switch	7	11	8

## D.1.2 Resultados medios obtenidos para cada nivel del dataset

Métrica	Nivel 1	Nivel 2	Nivel 3	Nivel 4	Nivel 5	Nivel 6
MOTP	0.73223	0.75224	0.70790	0.69915	0.70588	0.70832
MOTA	0.82776	0.70802	0.72665	0.59860	0.60656	0.56390
FP	0.10559	0.09427	0.01758	0.12926	0.09656	0.09241
TP	0.93335	0.80230	0.74422	0.72785	0.70311	0.65631
FN	0.06665	0.19770	0.25441	0.26422	0.29506	0.32945
ID switch	0.00	0.00	0.80	1.67	1.00	8.67

## D.2. Asignación de $\sigma_x$ y $\sigma_y$ a un valor de 6 para las secuencias que pertenecen al dataset público PETS09

### D.2.1 Resultados obtenidos para cada una de las secuencias de los niveles del dataset

#### D.2.1.1 Nivel 1

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.77681	0.72713	0.69276
MOTA	0.89372	0.67905	0.91052
FP	0.01932	0.27703	0.02041
TP	0.91304	0.95608	0.93093
FN	0.08696	0.04392	0.06907
ID switch	0	0	0

#### D.2.1.2 Nivel 2

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.76801	0.72279	0.76283
MOTA	0.61233	0.80882	0.70292
FP	0.25110	0.00735	0.02435
TP	0.86344	0.81618	0.72727
FN	0.13656	0.18382	0.27273
ID switch	0	0	0

#### D.2.1.3 Nivel 3

Métrica	Secuencia 1	Secuencia 2	Secuencia 3	Secuencia 4	Secuencia 5
MOTP	0.70110	0.68817	0.70339	0.74368	0.70774
MOTA	0.73083	0.59339	0.80101	0.86395	0.64407
FP	0.04211	0.00000	0.00000	0.00340	0.04237
TP	0.77293	0.59339	0.80101	0.86735	0.68644
FN	0.22556	0.40467	0.19899	0.12925	0.31356
ID switch	1	2	0	1	0

#### D.2.1.4 Nivel 4

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.68894	0.72625	0.68226
MOTA	0.81373	0.63514	0.34694
FP	0.01961	0.13007	0.23810
TP	0.83333	0.76520	0.58503
FN	0.16667	0.23142	0.39456
ID switch	0	2	3

#### D.2.1.5 Nivel 5

Métrica	Secuencia 1	Secuencia 2
MOTP	0.70010	0.71139
MOTA	0.71707	0.51743
FP	0.04878	0.13394
TP	0.76585	0.65138
FN	0.23415	0.34495
ID switch	0	2

#### D.2.1.6 Nivel 6

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.72066	0.69529	0.70714
MOTA	0.68142	0.58319	0.47111
FP	0.07257	0.12437	0.06815
TP	0.75398	0.70756	0.53926
FN	0.23540	0.27395	0.45185
ID switch	6	11	6

### D.2.1 Resultados medios obtenidos para cada nivel del dataset

Métrica	Nivel 1	Nivel 2	Nivel 3	Nivel 4	Nivel 5	Nivel 6
MOTP	0.73223	0.75121	0.70882	0.69915	0.70575	0.70770
MOTA	0.82776	0.70802	0.72665	0.59860	0.61725	0.57857
FP	0.10559	0.09427	0.01758	0.12926	0.09136	0.08836
TP	0.93335	0.80230	0.01758	0.72785	0.70862	0.66693
FN	0.06665	0.19770	0.74422	0.26422	0.28955	0.3204
ID switch	0.00	0.00	0.80	1.67	1.00	7.67

### D.3. Asignación de $\sigma_x$ y $\sigma_y$ a un valor de 7 para las secuencias que pertenecen al dataset público PETS09

#### D.3.1 Resultados obtenidos para cada una de las secuencias de los niveles del dataset

##### D.3.1.1 Nivel 1

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.77681	0.72713	0.69276
MOTA	0.89372	0.67905	0.91052
FP	0.01932	0.27703	0.02041
TP	0.91304	0.95608	0.93093
FN	0.08696	0.04392	0.06907
ID switch	0	0	0

##### D.3.1.2 Nivel 2

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.76801	0.72515	0.76283
MOTA	0.61233	0.80882	0.70292
FP	0.25110	0.00735	0.02435
TP	0.86344	0.81618	0.72727
FN	0.13656	0.18382	0.27273
ID switch	0	0	0

##### D.3.1.3 Nivel 3

Métrica	Secuencia 1	Secuencia 2	Secuencia 3	Secuencia 4	Secuencia 5
MOTP	0.70110	0.68817	0.70339	0.74368	0.71197
MOTA	0.73083	0.59339	0.80101	0.86395	0.64407
FP	0.04211	0.00000	0.00000	0.00340	0.04237
TP	0.77293	0.59339	0.80101	0.86735	0.68644
FN	0.22556	0.40467	0.19899	0.12925	0.31356
ID switch	1	2	0	1	0

##### D.3.1.4 Nivel 4

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.68894	0.72625	0.68226
MOTA	0.81373	0.63514	0.34694
FP	0.01961	0.13007	0.23810
TP	0.83333	0.76520	0.58503
FN	0.16667	0.23142	0.39456
ID switch	0	2	3

##### D.3.1.5 Nivel 5

Métrica	Secuencia 1	Secuencia 2
MOTP	0.70305	0.71178
MOTA	0.71220	0.52477
FP	0.05366	0.13028
TP	0.76585	0.65505
FN	0.23415	0.34128
ID switch	0	2

##### D.3.1.6 Nivel 6

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.72141	0.69556	0.70983
MOTA	0.67080	0.60672	0.46667
FP	0.08319	0.11597	0.06667
TP	0.75398	0.72269	0.53333
FN	0.23363	0.25882	0.45630
ID switch	7	11	7

### D.3.2 Resultados medios obtenidos para cada nivel del dataset

Métrica	Nivel 1	Nivel 2	Nivel 3	Nivel 4	Nivel 5	Nivel 6
MOTP	0.73223	0.75200	0.70966	0.69915	0.70742	0.70893
MOTA	0.82776	0.70802	0.72665	0.59860	0.61849	0.58140
FP	0.10559	0.09427	0.01758	0.12926	0.09197	0.08861
TP	0.93335	0.80230	0.74422	0.72785	0.71045	0.67000
FN	0.0667	0.19770	0.25441	0.26422	0.28772	0.31625
ID switch	0.00	0.00	0.80	1.67	1.00	8.33

## D.4. Asignación de $\sigma_x$ y $\sigma_y$ a un valor de 8 para las secuencias que pertenecen al *dataset* público PETS09

### D.4.1 Resultados obtenidos para cada una de las secuencias de los niveles del dataset

#### D.4.1.1 Nivel 1

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.77681	0.72713	0.69276
MOTA	0.89372	0.67905	0.91052
FP	0.01932	0.27703	0.02041
TP	0.91304	0.95608	0.93093
FN	0.08696	0.04392	0.06907
ID switch	0	0	0

#### D.4.1.1 Nivel 2

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.76801	0.72475	0.76283
MOTA	0.61233	0.80882	0.70292
FP	0.25110	0.00735	0.02435
TP	0.86344	0.81618	0.72727
FN	0.13656	0.18382	0.27273
ID switch	0	0	0

#### D.4.1.1 Nivel 3

Métrica	Secuencia 1	Secuencia 2	Secuencia 3	Secuencia 4	Secuencia 5
MOTP	0.70110	0.68817	0.70339	0.74368	0.70994
MOTA	0.73083	0.59339	0.80101	0.86395	0.64407
FP	0.04211	0.00000	0.00000	0.00340	0.04237
TP	0.77293	0.59339	0.80101	0.86735	0.68644
FN	0.22556	0.40467	0.19899	0.12925	0.31356
ID switch	1	2	0	1	0

#### D.4.1.1 Nivel 4

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.68894	0.72625	0.68226
MOTA	0.81373	0.63514	0.34694
FP	0.01961	0.13007	0.23810
TP	0.83333	0.76520	0.58503
FN	0.16667	0.23142	0.39456
ID switch	0	2	3

#### D.4.1.1 Nivel 5

Métrica	Secuencia 1	Secuencia 2
MOTP	0.70284	0.71400
MOTA	0.72683	0.53211
FP	0.04390	0.11927
TP	0.77073	0.65138
FN	0.22927	0.34495
ID switch	0	2

#### D.4.1.1 Nivel 6

Métrica	Secuencia 1	Secuencia 2	Secuencia 3
MOTP	0.72512	0.69348	0.70834
MOTA	0.65310	0.60504	0.46074
FP	0.07787	0.10756	0.06963
TP	0.73097	0.71261	0.53037
FN	0.24956	0.27059	0.45926
ID switch	11	10	7

### D.4.2 Resultados medios obtenidos para cada nivel del dataset

Métrica	Nivel 1	Nivel 2	Nivel 3	Nivel 4	Nivel 5	Nivel 6
MOTP	0.73223	0.75186	0.70926	0.69915	0.70842	0.70898
MOTA	0.82776	0.70802	0.72665	0.59860	0.62947	0.57296
FP	0.10559	0.09427	0.01758	0.12926	0.08159	0.08502
TP	0.93335	0.80230	0.74422	0.72785	0.71106	0.65798
FN	0.06665	0.19770	0.25441	0.26422	0.28711	0.32647
ID switch	0.00	0.00	0.80	1.67	1.00	9.33

### D.5. Comparativa de resultados medios obtenidos en función de los valores de $\sigma_x$ y $\sigma_y$

Métrica	Nivel 1				Nivel 2			
	$\sigma_x = 5$	$\sigma_x = 6$	$\sigma_x = 7$	$\sigma_x = 8$	$\sigma_x = 5$	$\sigma_x = 6$	$\sigma_x = 7$	$\sigma_x = 8$
	$\sigma_y = 5$	$\sigma_y = 6$	$\sigma_y = 7$	$\sigma_y = 8$	$\sigma_y = 5$	$\sigma_y = 6$	$\sigma_y = 7$	$\sigma_y = 8$
MOTP	0.73223	0.73223	0.73223	0.73223	0.75224	0.75121	0.75200	0.75186
MOTA	0.82776	0.82776	0.82776	0.82776	0.70802	0.70802	0.70802	0.70802
FP	0.10559	0.10559	0.10559	0.10559	0.09427	0.09427	0.09427	0.09427
TP	0.93335	0.93335	0.93335	0.93335	0.80230	0.80230	0.80230	0.80230
FN	0.06665	0.06665	0.06667	0.06665	0.19770	0.19770	0.19770	0.19770
ID switch	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Métrica	Nivel 3				Nivel 4			
	$\sigma_x = 5$	$\sigma_x = 6$	$\sigma_x = 7$	$\sigma_x = 8$	$\sigma_x = 5$	$\sigma_x = 6$	$\sigma_x = 7$	$\sigma_x = 8$
	$\sigma_y = 5$	$\sigma_y = 6$	$\sigma_y = 7$	$\sigma_y = 8$	$\sigma_y = 5$	$\sigma_y = 6$	$\sigma_y = 7$	$\sigma_y = 8$
MOTP	0.70790	0.70882	0.70966	0.70926	0.69915	0.69915	0.69915	0.69915
MOTA	0.72665	0.72665	0.72665	0.72665	0.59860	0.59860	0.59860	0.59860
FP	0.01758	0.01758	0.01758	0.01758	0.12926	0.12926	0.12926	0.12926
TP	0.74422	0.01758	0.74422	0.74422	0.72785	0.72785	0.72785	0.72785
FN	0.25441	0.74422	0.25441	0.25441	0.26422	0.26422	0.26422	0.26422
ID switch	0.80	0.80	0.80	0.80	1.67	1.67	1.67	1.67

Métrica	Nivel 5				Nivel 6			
	$\sigma_x = 5$	$\sigma_x = 6$	$\sigma_x = 7$	$\sigma_x = 8$	$\sigma_x = 5$	$\sigma_x = 6$	$\sigma_x = 7$	$\sigma_x = 8$
	$\sigma_y = 5$	$\sigma_y = 6$	$\sigma_y = 7$	$\sigma_y = 8$	$\sigma_y = 5$	$\sigma_y = 6$	$\sigma_y = 7$	$\sigma_y = 8$
MOTP	0.70588	0.70575	0.70742	0.70842	0.70832	0.70770	0.70893	0.70898
MOTA	0.60656	0.61725	0.61849	0.62947	0.56390	0.57857	0.58140	0.57296
FP	0.09656	0.09136	0.09197	0.08159	0.09241	0.08836	0.08861	0.08502
TP	0.70311	0.70862	0.71045	0.71106	0.65631	0.66693	0.67000	0.65798
FN	0.29506	0.28955	0.28772	0.28711	0.32945	0.3204	0.31625	0.32647
ID switch	1.00	1.00	1.00	1.00	8.67	7.67	8.33	9.33



## PRESUPUESTO

### 1) Ejecución Material

- Compra de ordenador personal (Software incluido)..... 2.000 €
- Material de oficina ..... 200 €
- Total de ejecución material ..... 2.200 €

### 2) Gastos generales

- 16 % sobre Ejecución Material ..... 352 €

### 3) Beneficio Industrial

- 6 % sobre Ejecución Material ..... 132 €

### 4) Honorarios Proyecto

- 640 horas a 15 € / hora..... 9600 €

### 5) Material fungible

- Gastos de impresión..... 60 €
- Encuadernación..... 200 €

### 6) Subtotal del presupuesto

- Subtotal Presupuesto..... 12060 €

### 7) I.V.A. aplicable

- 21% Subtotal Presupuesto ..... 2532,6 €

### 8) Total presupuesto

- Total Presupuesto..... 14593,6 €

Madrid, Junio de 2014

El Ingeniero Jefe de Proyecto

Fdo.: Raúl Porras Martín  
Ingeniero Superior de Telecomunicación



## **PLIEGO DE CONDICIONES**

Este documento contiene las condiciones legales que guiarán la realización, en este proyecto, de un SISTEMA DE SEGUIMIENTO DE PERSONAS EN VÍDEO BASADO EN DETECCIÓN. En lo que sigue, se supondrá que el proyecto ha sido encargado por una empresa cliente a una empresa consultora con la finalidad de realizar dicho sistema. Dicha empresa ha debido desarrollar una línea de investigación con objeto de elaborar el proyecto. Esta línea de investigación, junto con el posterior desarrollo de los programas está amparada por las condiciones particulares del siguiente pliego.

Supuesto que la utilización industrial de los métodos recogidos en el presente proyecto ha sido decidida por parte de la empresa cliente o de otras, la obra a realizar se regulará por las siguientes:

### **Condiciones generales**

1. La modalidad de contratación será el concurso. La adjudicación se hará, por tanto, a la proposición más favorable sin atender exclusivamente al valor económico, dependiendo de las mayores garantías ofrecidas. La empresa que somete el proyecto a concurso se reserva el derecho a declararlo desierto.

2. El montaje y mecanización completa de los equipos que intervengan será realizado totalmente por la empresa licitadora.

3. En la oferta, se hará constar el precio total por el que se compromete a realizar la obra y el tanto por ciento de baja que supone este precio en relación con un importe límite si este se hubiera fijado.

4. La obra se realizará bajo la dirección técnica de un Ingeniero Superior de Telecomunicación, auxiliado por el número de Ingenieros Técnicos y Programadores que se estime preciso para el desarrollo de la misma.

5. Aparte del Ingeniero Director, el contratista tendrá derecho a contratar al resto del personal, pudiendo ceder esta prerrogativa a favor del Ingeniero Director, quien no estará obligado a aceptarla.

6. El contratista tiene derecho a sacar copias a su costa de los planos, pliego de condiciones y presupuestos. El Ingeniero autor del proyecto autorizará con su firma las copias solicitadas por el contratista después de confrontarlas.

7. Se abonará al contratista la obra que realmente ejecute con sujeción al proyecto que sirvió de base para la contratación, a las modificaciones autorizadas por la superioridad o a las órdenes que con arreglo a sus facultades le hayan comunicado por escrito al Ingeniero Director de obras siempre que dicha obra se haya ajustado a los preceptos de los pliegos de condiciones, con arreglo a los cuales, se harán las modificaciones y la valoración de las diversas unidades sin que el importe total pueda exceder de los presupuestos aprobados. Por consiguiente, el número de unidades que se consignan en el proyecto o en el presupuesto, no podrá servirle de fundamento para entablar reclamaciones de ninguna clase, salvo en los casos de rescisión.

8. Tanto en las certificaciones de obras como en la liquidación final, se abonarán los trabajos realizados por el contratista a los precios de ejecución material que figuran en el presupuesto para cada unidad de la obra.

9. Si excepcionalmente se hubiera ejecutado algún trabajo que no se ajustase a las condiciones de la contrata pero que sin embargo es admisible a juicio del Ingeniero Director de obras, se dará conocimiento a la Dirección, proponiendo a la vez la rebaja de precios que el Ingeniero estime justa y si la Dirección resolviera aceptar la obra, quedará el contratista obligado a conformarse con la rebaja acordada.

10. Cuando se juzgue necesario emplear materiales o ejecutar obras que no figuren en el presupuesto de la contrata, se evaluará su importe a los precios asignados a otras obras o materiales análogos si los hubiere y cuando no, se discutirán entre el Ingeniero Director y el contratista, sometiéndolos a la aprobación de la Dirección. Los nuevos precios convenidos por uno u otro procedimiento, se sujetarán siempre al establecido en el punto anterior.

11. Cuando el contratista, con autorización del Ingeniero Director de obras, emplee materiales de calidad más elevada o de mayores dimensiones de lo estipulado en el proyecto, o sustituya una clase de fabricación por otra que tenga asignado mayor precio o ejecute con mayores dimensiones cualquier otra parte de las obras, o en general, introduzca en ellas cualquier modificación que sea beneficiosa a juicio del Ingeniero Director de obras, no tendrá derecho sin embargo, sino a lo que le correspondería si hubiera realizado la obra con estricta sujeción a lo proyectado y contratado.

12. Las cantidades calculadas para obras accesorias, aunque figuren por partidaalzada en el presupuesto final (general), no serán abonadas sino a los precios de la contrata, según las condiciones de la misma y los proyectos particulares que para ellas se formen, o en su defecto, por lo que resulte de su medición final.

13. El contratista queda obligado a abonar al Ingeniero autor del proyecto y director de obras así como a los Ingenieros Técnicos, el importe de sus respectivos honorarios facultativos por formación del proyecto, dirección técnica y administración en su caso, con arreglo a las tarifas y honorarios vigentes.

14. Concluida la ejecución de la obra, será reconocida por el Ingeniero Director que a tal efecto designe la empresa.

15. La garantía definitiva será del 4% del presupuesto y la provisional del 2%.

16. La forma de pago será por certificaciones mensuales de la obra ejecutada, de acuerdo con los precios del presupuesto, deducida la baja si la hubiera.

17. La fecha de comienzo de las obras será a partir de los 15 días naturales del replanteo oficial de las mismas y la definitiva, al año de haber ejecutado la provisional, procediéndose si no existe reclamación alguna, a la reclamación de la fianza.

18. Si el contratista al efectuar el replanteo, observase algún error en el proyecto, deberá comunicarlo en el plazo de quince días al Ingeniero Director de obras, pues transcurrido ese plazo será responsable de la exactitud del proyecto.

19. El contratista está obligado a designar una persona responsable que se entenderá con el Ingeniero Director de obras, o con el delegado que éste designe, para todo relacionado con ella. Al ser el Ingeniero Director de obras el que interpreta el proyecto, el contratista deberá consultarle cualquier duda que surja en su realización.

20. Durante la realización de la obra, se girarán visitas de inspección por personal facultativo de la empresa cliente, para hacer las comprobaciones que se crean oportunas. Es

obligación del contratista, la conservación de la obra ya ejecutada hasta la recepción de la misma, por lo que el deterioro parcial o total de ella, aunque sea por agentes atmosféricos u otras causas, deberá ser reparado o reconstruido por su cuenta.

21. El contratista, deberá realizar la obra en el plazo mencionado a partir de la fecha del contrato, incurriendo en multa, por retraso de la ejecución siempre que éste no sea debido a causas de fuerza mayor. A la terminación de la obra, se hará una recepción provisional previo reconocimiento y examen por la dirección técnica, el depositario de efectos, el interventor y el jefe de servicio o un representante, estampando su conformidad el contratista.

22. Hecha la recepción provisional, se certificará al contratista el resto de la obra, reservándose la administración el importe de los gastos de conservación de la misma hasta su recepción definitiva y la fianza durante el tiempo señalado como plazo de garantía. La recepción definitiva se hará en las mismas condiciones que la provisional, extendiéndose el acta correspondiente. El Director Técnico propondrá a la Junta Económica la devolución de la fianza al contratista de acuerdo con las condiciones económicas legales establecidas.

23. Las tarifas para la determinación de honorarios, reguladas por orden de la Presidencia del Gobierno el 19 de Octubre de 1961, se aplicarán sobre el denominado en la actualidad "Presupuesto de Ejecución de Contrata" y anteriormente llamado "Presupuesto de Ejecución Material" que hoy designa otro concepto.

### **Condiciones particulares**

La empresa consultora, que ha desarrollado el presente proyecto, lo entregará a la empresa cliente bajo las condiciones generales ya formuladas, debiendo añadirse las siguientes condiciones particulares:

1. La propiedad intelectual de los procesos descritos y analizados en el presente trabajo, pertenece por entero a la empresa consultora representada por el Ingeniero Director del Proyecto.
2. La empresa consultora se reserva el derecho a la utilización total o parcial de los resultados de la investigación realizada para desarrollar el siguiente proyecto, bien para su publicación o bien para su uso en trabajos o proyectos posteriores, para la misma empresa cliente o para otra.
3. Cualquier tipo de reproducción aparte de las reseñadas en las condiciones generales, bien sea para uso particular de la empresa cliente, o para cualquier otra aplicación, contará con autorización expresa y por escrito del Ingeniero Director del Proyecto, que actuará en representación de la empresa consultora.
4. En la autorización se ha de hacer constar la aplicación a que se destinan sus reproducciones así como su cantidad.
5. En todas las reproducciones se indicará su procedencia, explicitando el nombre del proyecto, nombre del Ingeniero Director y de la empresa consultora.
6. Si el proyecto pasa la etapa de desarrollo, cualquier modificación que se realice sobre él, deberá ser notificada al Ingeniero Director del Proyecto y a criterio de éste, la empresa consultora decidirá aceptar o no la modificación propuesta.

7. Si la modificación se acepta, la empresa consultora se hará responsable al mismo nivel que el proyecto inicial del que resulta el añadirla.

8. Si la modificación no es aceptada, por el contrario, la empresa consultora declinará toda responsabilidad que se derive de la aplicación o influencia de la misma.

9. Si la empresa cliente decide desarrollar industrialmente uno o varios productos en los que resulte parcial o totalmente aplicable el estudio de este proyecto, deberá comunicarlo a la empresa consultora.

10. La empresa consultora no se responsabiliza de los efectos laterales que se puedan producir en el momento en que se utilice la herramienta objeto del presente proyecto para la realización de otras aplicaciones.

11. La empresa consultora tendrá prioridad respecto a otras en la elaboración de los proyectos auxiliares que fuese necesario desarrollar para dicha aplicación industrial, siempre que no haga explícita renuncia a este hecho. En este caso, deberá autorizar expresamente los proyectos presentados por otros.

12. El Ingeniero Director del presente proyecto, será el responsable de la dirección de la aplicación industrial siempre que la empresa consultora lo estime oportuno. En caso contrario, la persona designada deberá contar con la autorización del mismo, quien delegará en él las responsabilidades que ostente.