UNIVERSIDAD AUTONOMA DE MADRID

ESCUELA POLITECNICA SUPERIOR

PROYECTO FIN DE CARRERA

Segmentación de secuencias de vídeo basada en el modelado del fondo mediante capas.

Alfonso Colmenarejo Rubio
Julio 2011

Segmentación de secuencias de vídeo basada en el modelado del fondo mediante capas.

AUTOR: Alfonso Colmenarejo Rubio

Tutor: Marcos Escudero Viñolo

Ponente: Jesús Bescós Cano

Video Processing and UnderstandingLab

Dpto. de Ingeniería Informática

Escuela Politécnica Superior

Universidad Autónoma de Madrid

Julio de 2011

Palabras clave

Segmentación, fondo, frente, modelado de fondo, modelado de frente, sustracción de fondo, Gaussiana Simple (*SG*), Mezcla de Gaussianas (*MoG*), Bayesiano, clasificación, fondo estático (unimodal), fondo dinámico(multimodal).

Resumen

El principal objetivo de este proyecto es el diseño, implementación y mejora progresiva de un algoritmo de segmentación de objetos por sustracción del fondo, que modele las diferentes apariencias que puede adoptar un pixel a lo largo de un vídeo en capas independientes. Los sistemas existentes en el modelado multicapa del fondo suelen utilizar aproximaciones Bayesianas para la actualización del modelo y aplican decisiones a posteriori sobre la clase a la que pertenece el pixel (fondo, frente, sombras, etc.). La principal aportación al estado del arte de este proyecto es la utilización de un esquema de clases previo al modelado del fondo. Mediante esta clasificación a priori en diferentes clases de fondo se evitará la introducción de muestras del frente en el modelo que, aunque pueden ser aisladas en el proceso de decisión a posteriori de cada cuadro, influirán en las decisiones de actualización y discriminación en cuadros posteriores.

El proyecto se ha desarrollado de manera incremental, partiendo de la implementación de un algoritmo base e introduciendo sistemas que incluyen soluciones a los problemas teóricos del sistema anterior. Los resultados obtenidos muestran la eficacia de cada una de las mejoras y la robustez del sistema final a fondos dinámicos, inicios en caliente y cambios bruscos de iluminación.

Abstract

The main purpose of the proposed project is the design, implementation and progressive improvement of a background subtraction video segmentation system that works by modeling the different appearances of a pixel in a set of independent layers. State of the Art systems in the area of multilayer background modeling make use of Bayesian based schemes for the process of model updating and perform a posteriori decisions to discriminate among possible pixel classes (background, foreground, shadows, etc.). The main contribution of this project to the existing approaches is the use of an a priori classification scheme that assigns the pixel a set of predefined background classes. By means of this classification, which is performed previously to the model update, the algorithm is capable to isolate the pixel foreground samples and to avoid its influence in the updating and discrimination processes performed over the subsequent frames.

The project has been developed hierarchically, starting from the description and implementation of a Bayesian base algorithm and proposing new schemes solving theoretical problems of the predecessor systems. Obtained results demonstrate the effectiveness of each of the proposed improvements and the adequate performance of the final system in the presence of highly dynamic backgrounds, hot starts and global illumination changes

Agradecimientos

En primer lugar quiero dar las gracias a mi tutor, Marcos Escudero Viñolo, por su apoyo y dedicación a lo largo de este proyecto. También quiero agradecer a Jesús Bescós y a José María Martínez su confianza al permitirme la realización de este Proyecto Fin de Carrera en el VPULab y a todos los integrantes del laboratorio por los buenos ratos que hemos pasado en él.

Gracias a mi padre Mariano, a mi madre MªAngela, a mi hermana María y al resto de mi familia por su apoyo a lo largo de toda mi vida en los buenos y en los malos momentos que me ha permitido llegar hasta aquí.

También quiero darles las gracias a todos los amigos que he conocido a lo largo de estos difíciles pero intensos y divertidos años de vida universitaria. Por los buenos momentos que hemos pasado en la universidad y fuera de ella y por los buenos momentos que seguiremos pasando una vez acabados los años de vida universitaria.

Por último quiero dar las gracias a Laura por su apoyo y cariño durante todo este tiempo ya que sin ti estos años en la universidad no hubieran sido lo mismo. Espero que en el futuro todo nos valla igual de bien que hasta ahora y que emprendamos muchos proyectos juntos.

Alfonso Colmenarejo Rubio.

INDICE DE CONTENIDOS.

1 Introducción. 1
1.1 Motivación. 1
1.2 Objetivos. 2
1.3 Organización de la memoria. 4
2 Nomenclatura utilizada. 5
3 Estado del arte. 9
3.1 Introducción a la segmentación frente/fondo. 9
3.2 El pixel como unidad de análisis. 10
3.3 Problemas del modelado de fondo. 11
3.4 La segmentación como un problema de clasificación. 12
3.5 Modelos de sustracción del fondo. 20
3.5.1 Esquema general de sustracción del fondo. (BS) 20
3.5.2 Modelado de fondo. 21
3.5.3 Técnicas concretas de modelado no Bayesianas. 23
3.5.3.1 Un modo. Fondo unimodal. 23
3.5.3.2 Dos o más modos. Fondo multimodal. 24
3.5.4 Modelado del fondo por esquema Bayesianos multicapa. 26
3.5.4.1 Ejemplos de técnicas de modelado Bayesiano Multicapa. 30
3.5.4.2 Comparativa Bayesiano vs Mezcla de Gaussianas. 31
3.5.5 Discriminación frente fondo. 31
3.5.5.1 Similitud al modelo. 32
3.6 Tratamiento a nivel de blob. 33
3.7 Modelado de frente. 35
4 Procesamiento interno en cada capa. 37
4.1 Parámetros internos a la capa. 37
4.2 Cálculo de la probabilidad de pertenencia a una capa: $p\left(z_t \mid \vec{x}_t\right)$ 37
4.3 Distancia de Mahalanobis. 38
4.3.1 Modelado de la distancia. 39
4.3.2 Adaptación de parámetros. 39
4.4 Umbralización automática de la distancia. 40
5 Diseño básico. Modelado Bayesiano del Fondo. 43
5.1 Cálculo de la probabilidad de pertenencia al modelo de fondo: $p\left(BG_t \mid z_t, \vec{x}_t\right)$ 43

5.2 Arquitectura del sistema. 43
5.2.1 Descripción de la arquitectura del sistema. 44
5.3 Modelado de fondo. 46
5.3.1 Actualización de la confianza 47
5.4 La confianza como parámetro de discriminación frente-fondo. 48
5.5 Conclusiones y limitaciones del sistema. 48
6 Descripción de las mejoras. 51
6.1 Presentación de las mejoras introducidas. 51
6.2 Introducción de la matriz de covarianza completa. 51
6.3 Modelado de fondo utilizando clasificación de píxel. 52
6.3.1 Clases de pixel. 54
6.3.2 Utilizando información de bajo nivel (pixel aislado). 55
6.3.2.1 Arquitectura del sistema. 56
6.3.2.2 Descripción de la arquitectura del sistema. 56
6.3.2.3 Resumen de las características analizadas. 60
6.3.3 Utilizando información a nivel de blob. 61
6.3.3.1 Arquitectura del sistema. 62
6.3.3.2 Descripción de la arquitectura del sistema. 62
6.4 Modelado de frente. 68
6.4.1 Arquitectura del sistema. 69
6.4.1.1 Descripción de la arquitectura del sistema. 69
6.4.1.2 Limitaciones del sistema final. 71
6.5 Resumen de las mejoras y sistemas implementados para su evaluación. 72
7 Resultados. 75
7.1 Descripción de las secuencias de prueba. 75
7.2 .Configuración inicial del sistema. 81
7.3 Análisis comparativo de los sistemas implementados. 84
7.3.1 Coste computacional de cada uno de los sistemas. 84
7.3.2 Métricas utilizadas. 88
7.3.3 Calidad de la segmentación. 90
7.3.4 Influencia de la inicialización de σ_0 134
7.3.5 Comparativa con el SoA. 136
7.4 Problemas solucionados. 138

7.5 Resultados en proceso de publicación e integración en sistema comercial. 139
8 Conclusiones y trabajo futuro. 141
9 Referencias. 145
Anexos. 149
Anexo A: desarrollo del modelado de cada capa. 149
Anexo B: cálculo de la inversa de la matriz de covarianza completa. 153
Anexo C: publicaciones. 155
PRESUPUESTO. 157

INDICE DE FIGURAS

FIGURA 3-1: EVOLUCIÓN DE *Y* EN UN PIXEL DE FONDO UNIMODAL. 13

FIGURA 3-2: EVOLUCIÓN DE *Y* EN UN PIXEL DE FONDO MULTIMODAL. 14

FIGURA 3-3: EVOLUCIÓN DE *Y* EN UN PIXEL DE FONDO UNIMODAL CON PRESENCIA DE FRENTE. 15

FIGURA 3-4: EVOLUCIÓN DE *Y* EN UN PIXEL DE FONDO MULTIMODAL CON PRESENCIA DE FRENTE. 16

FIGURA 3-5: EVOLUCIÓN DE *Y* EN PIXEL DE FONDO UNIMODAL CON PRESENCIA DE SOMBRAS.. 18

FIGURA 3-6: EVOLUCIÓN DE *Y* EN PIXEL FONDO UNIMODAL CON CAMBIOS BRUSCOS DE ILUMINACIÓN. 19

FIGURA 3-7: ESQUEMA DE SUSTRACCIÓN DE FONDO (*BS*). 20

FIGURA 3-8: EJEMPLO DE EXTRACCIÓN DE BLOBS. 35

FIGURA 4-1: MODELADO DE LA DISTANCIA. 41

FIGURA 5-1: ESQUEMA DEL SISTEMA *BAYESIANO BÁSICO*. 43

FIGURA 6-1: ESQUEMA DEL SISTEMA MEJORADO A NIVEL DE PIXEL. 56

FIGURA 6-2: VARIABILIDAD DE LA CONFIANZA. 58

FIGURA 6-3: ESQUEMA DEL SISTEMA MEJORADO A NIVEL DE BLOB. 62

FIGURA 6-4: CARACTERIZACIÓN DEL BLOB. 64

FIGURA 6-5: ESQUEMA DEL SISTEMA MEJORADO A NIVEL DE BLOB CON MODELADO DE FRENTE. 69

FIGURA 7-1: TIEMPOS MEDIOS DE PROCESAMIENTO DE LOS ALGORITMOS DESARROLLADOS. ... 87

FIGURA 7-2: DIAGRAMAS DE BARRAS DE LOS ESTADÍSTICOS DE S1. 92

FIGURA 7-3: DIAGRAMAS DE BARRAS DE LOS ESTADÍSTICOS DE S2. 95

FIGURA 7-4: DIAGRAMAS DE BARRAS DE LOS ESTADÍSTICOS DE S3. 98

FIGURA 7-5: DIAGRAMAS DE BARRAS DE LOS ESTADÍSTICOS DE S4. 101

FIGURA 7-6: DIAGRAMAS DE BARRAS DE LOS ESTADÍSTICOS DE S5. 104

FIGURA 7-7: DIAGRAMAS DE BARRAS DE LOS ESTADÍSTICOS DE S6. 107

FIGURA 7-8: DIAGRAMAS DE BARRAS DE LOS ESTADÍSTICOS DE S7. 110

FIGURA 7-9: DIAGRAMAS DE BARRAS DE LOS ESTADÍSTICOS DE S8. 114

FIGURA 7-10: DIAGRAMAS DE BARRAS DE LOS ESTADÍSTICOS DE S9. 117

FIGURA 7-11: DIAGRAMAS DE BARRAS DE LOS ESTADÍSTICOS DE S10. 120

FIGURA 7-12: DIAGRAMAS DE BARRAS DE LOS ESTADÍSTICOS DE S11. 123

FIGURA 7-13: DIAGRAMAS DE BARRAS DE LOS ESTADÍSTICOS DE S12. 126

FIGURA 7-14: DIAGRAMAS DE BARRAS DE LOS ESTADÍSTICOS DE S13. 129

FIGURA 7-15: DIAGRAMAS DE BARRAS DE LOS ESTADÍSTICOS DE S14. 132

FIGURA 7-16: CURVAS ROC PARA EVALUAR LA INICIALIZACIÓN DE σ_0 136

FIGURA 7-17: RESULTADOS CUALITATIVOS Y PROBLEMAS SOLVENTADOS. 139

INDICE DE TABLAS

TABLA 6-1. CLASIFICACIÓN DE PIXEL EN FUNCIÓN DE LA VARIABILIDAD DE LA CONFIANZA.... 60

TABLA 6-2. RESUMEN DE LAS MEJORAS Y SISTEMAS IMPLEMENTADOS. 73

TABLA 7-1. SECUENCIAS DE PRUEBA. 79

TABLA 7-2. CARACTERÍSTICAS DE LAS SECUENCIAS DE PRUEBA. 80

TABLA 7-3. CLASIFICACIÓN DE PIXEL EN FUNCIÓN DE LA VARIABILIDAD DE LA CONFIANZA.... 83

TABLA 7-4. TIEMPOS DE PROCESAMIENTO UTILIZANDO MODELADO CON K=3 PARA EL MODELO DE FONDO. 85

TABLA 7-5. TIEMPOS DE PROCESAMIENTO UTILIZANDO MODELADO CON K=5 MÁXIMAS PARA EL FONDO. 86

TABLA 7-6. PARÁMETROS ESTADÍSTICOS. 88

TABLA 7-7. RESULTADOS COMPARATIVOS SOBRE LA SECUENCIA S1 CON MODELO DE FONDO DE 3 CAPAS. 91

TABLA 7-8. RESULTADOS COMPARATIVOS SOBRE LA SECUENCIA S1 CON MODELO DE FONDO DE 5 CAPAS. 92

TABLA 7-9. RESULTADOS CUALITATIVOS SOBRE LA SECUENCIA S1. 93

TABLA 7-10. RESULTADOS COMPARATIVOS SOBRE LA SECUENCIA S2 CON MODELO DE FONDO DE 3 CAPAS 94

TABLA 7-11. RESULTADOS COMPARATIVOS SOBRE LA SECUENCIA S2 CON MODELO DE FONDO DE 5 CAPAS 95

TABLA 7-12. RESULTADOS CUALITATIVOS SOBRE LA SECUENCIA S2. 96

TABLA 7-13. RESULTADOS COMPARATIVOS SOBRE LA SECUENCIA S3 CON MODELO DE FONDO DE 3 CAPAS 97

TABLA 7-14. RESULTADOS COMPARATIVOS SOBRE LA SECUENCIA S3 CON MODELO DE FONDO DE 5 CAPAS 97

TABLA 7-15. RESULTADOS CUALITATIVOS SOBRE LA SECUENCIA S3. 99

TABLA 7-16. RESULTADOS COMPARATIVOS SOBRE LA SECUENCIA S4 CON MODELO DE FONDO DE 3 CAPAS 100

TABLA 7-17. RESULTADOS COMPARATIVOS SOBRE LA SECUENCIA S4 CON MODELO DE FONDO DE 5 CAPAS 101

TABLA 7-18. RESULTADOS CUALITATIVOS SOBRE LA SECUENCIA S4. 102

TABLA 7-19. RESULTADOS COMPARATIVOS SOBRE LA SECUENCIA S5 CON MODELO DE FONDO DE 3 CAPAS 103

TABLA 7-20. RESULTADOS COMPARATIVOS SOBRE LA SECUENCIA S5 CON MODELO DE FONDO DE 5 CAPAS 103

TABLA 7-21. RESULTADOS CUALITATIVOS SOBRE LA SECUENCIA S5. 105

TABLA 7-22. RESULTADOS COMPARATIVOS SOBRE LA SECUENCIA S6 CON MODELO DE FONDO DE 3 CAPAS 106

TABLA 7-23. RESULTADOS COMPARATIVOS SOBRE LA SECUENCIA S6 CON MODELO DE FONDO DE 5 CAPAS 107

TABLA 7-24. RESULTADOS CUALITATIVOS SOBRE LA SECUENCIA S6. 108

TABLA 7-25. RESULTADOS COMPARATIVOS SOBRE LA SECUENCIA S7 CON MODELO DE FONDO DE 3 CAPAS 109

TABLA 7-26. RESULTADOS COMPARATIVOS SOBRE LA SECUENCIA S7 CON MODELO DE FONDO DE 5 CAPAS..... 110

TABLA 7-27. RESULTADOS CUALITATIVOS SOBRE LA SECUENCIA S7. 111

TABLA 7-28. RESULTADOS COMPARATIVOS SOBRE LA SECUENCIA S8 CON MODELO DE FONDO DE 3 CAPAS..... 113

TABLA 7-29. RESULTADOS COMPARATIVOS SOBRE LA SECUENCIA S8 CON MODELO DE FONDO DE 5 CAPAS..... 113

TABLA 7-30. RESULTADOS CUALITATIVOS SOBRE LA SECUENCIA S8. 115

TABLA 7-31. RESULTADOS COMPARATIVOS SOBRE LA SECUENCIA S9 CON MODELO DE FONDO DE 3 CAPAS..... 116

TABLA 7-32. RESULTADOS COMPARATIVOS SOBRE LA SECUENCIA S9 CON MODELO DE FONDO DE 5 CAPAS..... 117

TABLA 7-33. RESULTADOS CUALITATIVOS SOBRE LA SECUENCIA S9. 118

TABLA 7-34. RESULTADOS COMPARATIVOS SOBRE LA SECUENCIA S10 CON MODELO DE FONDO DE 3 CAPAS..... 119

TABLA 7-35. RESULTADOS COMPARATIVOS SOBRE LA SECUENCIA S10 CON MODELO DE FONDO DE 5 CAPAS..... 119

TABLA 7-36. RESULTADOS CUALITATIVOS SOBRE LA SECUENCIA S10. 121

TABLA 7-37. RESULTADOS COMPARATIVOS SOBRE LA SECUENCIA S11 CON MODELO DE FONDO DE 3 CAPAS..... 122

TABLA 7-38. RESULTADOS COMPARATIVOS SOBRE LA SECUENCIA S11 CON MODELO DE FONDO DE 5 CAPAS..... 123

TABLA 7-39. RESULTADOS CUALITATIVOS SOBRE LA SECUENCIA S11. 124

TABLA 7-40. RESULTADOS COMPARATIVOS SOBRE LA SECUENCIA S12 CON MODELO DE FONDO DE 3 CAPAS..... 125

TABLA 7-41. RESULTADOS COMPARATIVOS SOBRE LA SECUENCIA S12 CON MODELO DE FONDO DE 5 CAPAS..... 126

TABLA 7-42. RESULTADOS CUALITATIVOS SOBRE LA SECUENCIA S12. 127

TABLA 7-43. RESULTADOS COMPARATIVOS SOBRE LA SECUENCIA S13 CON MODELO DE FONDO DE 3 CAPAS..... 128

TABLA 7-44. RESULTADOS COMPARATIVOS SOBRE LA SECUENCIA S13 CON MODELO DE FONDO DE 5 CAPAS..... 128

TABLA 7-45. RESULTADOS CUALITATIVOS SOBRE LA SECUENCIA S13. 130

TABLA 7-46. RESULTADOS COMPARATIVOS SOBRE LA SECUENCIA S14 CON MODELO DE FONDO DE 3 CAPAS..... 131

TABLA 7-47. RESULTADOS COMPARATIVOS SOBRE LA SECUENCIA S14 CON MODELO DE FONDO DE 5 CAPAS..... 131

TABLA 7-48. RESULTADOS CUALITATIVOS SOBRE LA SECUENCIA S14. 133

TABLA 7-49. RESULTADOS COMPARATIVOS CON [2] SOBRE LAS SECUENCIAS S1, S6, S7, S8, S10 137

TABLA 7-50. RESULTADOS COMPARATIVOS CON [46] SOBRE LAS SECUENCIAS S5, S11 Y S14.138

TABLA 7-51. RESULTADOS COMPARATIVOS CON [46] SOBRE LA SECUENCIA S11 (INICIALIZACIÓN ÓPTIMA). 138

1 Introducción.

1.1 Motivación.

La segmentación de objetos en movimiento, o discriminación entre frente (*foreground*) y fondo (*background*), es una etapa clave en la mayoría de los sistemas de análisis de video. Si el objetivo último es ofrecer una descripción de la escena de alto nivel, indicando el número, la posición y el movimiento de objetos (animados e inanimados, rígidos o flexibles, etc.), y la interacción de éstos objetos entre sí, los resultados obtenidos tras la segmentación a nivel de píxel (también conocida como de bajo nivel) influirán irremediablemente en la calidad final del sistema.

Las técnicas clásicas de segmentación a nivel de píxel buscan modelar el fondo mediante una capa o imagen del fondo que tiene que ser actualizada con cada nuevo cuadro del vídeo. El frente se detecta como alteraciones sobre este modelo, técnica habitualmente conocida como substracción del fondo. Los resultados obtenidos por éstas técnicas son aceptables en entornos donde el fondo es simple y estático, la calidad del video es buena y la complejidad de los objetos baja.

Asumiendo que la cámara que graba la secuencia está fija, existen diversos métodos para llevar a cabo la substracción del fondo; el más sencillo consiste en hallar la diferencia entre la imagen actual y el modelo de fondo de referencia. Ahora bien, para discernir entre lo que es ruido en la imagen y lo que realmente se está moviendo, se requieren métodos que identifiquen qué píxeles de la imagen son objeto en movimiento y cuáles pertenecen al fondo. Para ello, existen modelos de fondo en los que a cada píxel se asocia un rango de valores probables de intensidad, por ejemplo a través de una distribución unimodal Gaussiana. De este modo, podemos desechar factores como cambios de iluminación progresivos en la escena, ruido inherente en el sistema de captación de vídeo, etc. No obstante, podemos encontrarnos ante otro tipo de variaciones del fondo, como cambios bruscos de iluminación o movimientos de las ramas de los árboles, arbustos, etc. De hecho, existe un sinfín de elementos en movimiento en las escenas de vídeo que no pueden modelarse a través de un único parámetro. La mezcla de Gaussianas, se utiliza para resolver esta situación: fondos complejos y no estáticos. Este modelo también posee

algunas desventajas ya que ante cambios rápidos en las escenas se incrementa el número de Gaussianas necesarias para modelarlos, causando problemas en la detección e incremento en el coste computacional del algoritmo.

Considerando estas limitaciones, en el contexto de este proyecto fin de carrera se han estudiado sistemas que enriquecen el modelado del fondo mediante la utilización de varios sub-modelos para cada píxel, exportando los resultados antes obtenidos a entornos donde el fondo no es estático, es decir, que contiene objetos en movimiento. Dentro de estos modelos, podríamos incluir el modelado Bayesiano multicapa [1], propuesto recientemente. El sistema propone modelar cada píxel del fondo como un conjunto de capas que compiten entre sí. Este modelado del fondo, permite no sólo estimar la distribución de los datos en cada capa, sino también la probabilidad de ocurrencia de cada una de estas distribuciones y utilizar estas probabilidades en el proceso de actualización del fondo y discriminación del frente.

1.2 **Objetivos.**

Los objetivos de este proyecto fin de carrera serán; estudio y análisis de técnicas de segmentación basadas en modelado de fondo [2] y [3], la implementación de un esquema de segmentación del frente mediante modelado Bayesiano [1], el diseño e implementación de mejoras sobre dicho sistema y la evaluación de los resultados obtenidos frente a otros sistemas disponibles en el VPU-Lab [2] u otros trabajos.

Debido a que el objetivo del proyecto es trabajar en entornos lo más genéricos posible (permaneciendo siempre en entornos de cámara fija o de movimiento de cámara compensando), las secuencias de video a analizar estarán sometidas a una serie de factores que desembocarán en problemas de diferentes complejidades que buscaremos solventar.

Entre estos factores de complejidad podemos mencionar:

- La obtención y actualización del fondo de la escena.

- La determinación de los parámetros de funcionamiento de los algoritmos para discriminar entre fondo y objetos en las diferentes secuencias (umbrales, pesos,…)

- La influencia del ruido introducido durante la captación de la secuencia realizada por la cámara.

- Los cambios de iluminación globales a los que puede estar sometida la escena a analizar y la velocidad de la transición de éstos (rápidos/lentos).

- Los cambios de iluminación locales, es decir que sólo afectan a una parte de la escena, entre los que incluiríamos las sombras y los reflejos producidos por la interacción fuentes de luz-objetos y fuentes de luz-fondo.

- Los fondos multimodales: aquellos en los que existen objetos en movimiento distintos a los que son objetivo del análisis (hojas de árboles bajo la influencia del viento, ondas en el agua, fuego…).

Adicionalmente se propondrán mejoras sobre los algoritmos implementados que permitan exportar su funcionamiento al modelado del frente [4] desarrollando sistemas similares a los más recientes en el área de segmentación a nivel de píxel o su utilización en situaciones donde los objetos objetivo del análisis estén presentes al comienzo del video (situaciones conocidas como de *'inicio en caliente'*).

Finalmente se estudiará la utilización de fuentes de información distintas al valor cromático de cada píxel (comúnmente *RGB*), como por ejemplo, la estimación del movimiento aparente al que están sometidos los objetos del frente, el solapamiento de objetos en imágenes consecutivas o parámetros internos del modelo como es el caso de la evolución de la medida de confianza en el modelo Bayesiano multicapa propuesto en [1].

1.3 Organización de la memoria.

La memoria consta de los siguientes capítulos:

Capítulo1: introducción, objetivos y motivación del proyecto.

Capítulo 2: nomenclatura básica.

Capítulo 3: estado del arte. Métodos de segmentación, modelado de fondo y modelado de frente.

Capítulo 4: procesamiento interno en cada capa.

Capítulo 5: diseño básico. Modelado Bayesiano del fondo.

Capítulo 6: descripción de las mejoras. Discriminación utilizando matriz de covarianza completa, y clasificación de pixel utilizando características a nivel de píxel, a nivel de blob, y modelo de frente.

Capítulo 7: pruebas y resultados.

Capítulo 8: conclusiones y trabajo futuro.

2 Nomenclatura utilizada.

t	Instante de tiempo.
N	Número de dimensiones de los datos de entrada.
x_t	Pixel (genérico)
$\vec{x}_t$	Pixel muestra de entrada (vector N dimensional)
(x, y)	Coordenada espacial del pixel.
BG	Modelo de fondo.
BG_t	Modelo de fondo en el instante de tiempo t.
FG	Modelo de frente.
FG_t	Modelo de frente en el instante de tiempo t.
Y	Luminancia.
RGB	Escala de color del sistema visual humano.
n	Número indefinido.
α	Peso de actualización.
μ	Valor medio de una distribución Gaussiana.
μ_t	Valor medio de una distribución en el instante t.
σ_t^2	Varianza de una distribución en el instante t.
σ_t	Desviación típica de una distribución en el instante t.
σ_Z	Desviación típica del canal Z (p.e. R,G,B)

SG	Gaussiana Simple.
MoG	Mezcla de Gaussianas.
modo	Cada una de las apariencias que puede adoptar un píxel a lo largo de un video.
capa	Contiene un modo y su descripción.
K	Número de capas máximas del modelo de fondo.
z	Variable discreta que representa el conjunto de las k capas de las que se compone el modelo de fondo.
w_k	Peso de cada Gaussiana.
$w_{k,t}$	Peso de cada Gaussiana en el instante t.
d	Distancia Euclídea.
D	Distancia Mahalanobis.
C	Medida de confianza.
Σ	Matriz de covarianza.
Σ_t	Matriz de covarianza en el intante t.
m_t	Medias almacenadas del modelo de fondo (genérica)
$\vec{m}_t$	Media del modelo de fondo (N dimensional)
K	Número de desviaciones típicas permitido para pertenecer a una distribución Gaussiana.
PoFG	Píxeles clasificados como frente.
PoBG	Píxeles clasificados como fondo.

I	Matriz Identidad.
v_t	Grados de libertad del modelo.
κ_t	Número de muestras utilizadas en el aprendizaje del modelo.
Λ_t	Matriz de escala del modelo.
BS	Algoritmo de sustracción de fondo. (Background Substraction)
X	Variable aleatoria (genérica).
U_C	Umbral de confianza para decidir si el modelo de fondo es fiable.
M	Variable aleatoria que representa la esperanza de X.
$BlobId$	Número con el que se identifica a los diferentes blobs que se detectan en el sistema.
ID	Identificador de las distintas secuencias de prueba.
GT	Máscara de *ground-truth* contra la que se comparan los resultados obtenidos.
$\begin{pmatrix} a & b & c \\ b & d & e \\ c & e & f \end{pmatrix}$	Posiciones dentro de una matriz 3x3.
SX	Secuencia numero X que se utiliza para evaluar los resultados.
β	Parámetro para inicializar σ_0

3 Estado del arte.

3.1 Introducción a la segmentación frente/fondo.

Debido a la creciente cantidad de información visual capturada por las cámaras y almacenada en servidores de internet, es necesario desarrollar herramientas de análisis automático en ciertos dominios de aplicación. En este contexto, el primer problema es la localización de la región donde sucede algo relevante. Esta operación suele conocerse como segmentación. Actualmente se está potenciando el estudio en profundidad de estas técnicas, ya que son un proceso fundamental sin el cuál sería imposible desarrollar sistemas de análisis de secuencias de vídeo de alto nivel como: video-vigilancia [5] y [6]; indexación de contenidos multimedia [7]; cine y TV [8]; o compresión de video[9].

El primer problema que surge en la segmentación de vídeo es el de definir el fondo de la escena (BG_t) El fondo de un video puede dividirse en volúmenes espacio-temporales que describen la evolución de una serie de características que lo describen a lo largo del tiempo. Estos volúmenes pueden ser espacialmente tan pequeños como el píxel (definido como x_t y posicionado en la coordenada (x, y) del cuadro) o tan grandes como la resolución total de cada cuadro del video. Temporalmente, se almacenarían las características del área espacial seleccionada a lo largo del video y, a partir de estas estadísticas se estimaría el modelo de fondo.

Definimos *fondo* como la parte de la imagen que es intrínseca a la escena grabada, es decir que permanece en la escena durante todo el vídeo. Puede ser una zona estática, que permanece constante durante todo el vídeo, también conocido como *fondo unimodal*, o puede ser un conjunto de zonas definidas por varios modos que se van alternando en escena. Estas zonas se denominan zonas de *fondo multimodal*.

Se definen como *fondo unimodal* todos aquellos píxeles que pueden describirse durante todo el vídeo por un único modo que se mantiene constante o que tiene cambios lentos. En este tipo de píxeles la obtención del modelo de fondo y la posterior

segmentación es más sencilla que en los fondos multimodales. Un ejemplo de fondo unimodal sería una pared blanca y lisa.

Se define como *fondo multimodal* todos los píxeles que no se pueden describir mediante un único modo. Esto es debido a que en ellos el fondo puede tener dos o más aspectos que pueden ser muy diferentes entre ellos. Un ejemplo de fondo multimodal serían las olas del mar que se mueven a causa del viento o la rama de un árbol que deja ver lo que tiene detrás de ella.

Definimos *frente* como las zonas de la imagen que no se adaptan al modelo de fondo. Va variando durante todo el vídeo y suelen ser las zonas que, dependiendo de la aplicación, interesa detectar para su posterior tratamiento.

3.2 **El pixel como unidad de análisis.**

Si aceptamos la división realizada en el proceso de captación de la escena realizado por la cámara, podemos utilizar el píxel como la unidad espacio-temporal básica. Cada píxel sería una muestra de la evolución temporal de las características que lo describen. Existen multitud de trabajos que utilizan el pixel como unidad de análisis para realizar algoritmos de segmentación de vídeo automática. Por ejemplo Piccardi en [3], hace una revisión de las técnicas más conocidas de segmentación de bajo nivel utilizando sustracción del fondo. Fatih Porikli en [1], propone un algoritmo de segmentación que utiliza el pixel como unidad de análisis pero utilizando un algoritmo de aprendizaje bayesiano para actualizar el modelo de fondo.

En estos trabajos, se define un modelo de fondo describiendo cada uno de los píxeles de la imagen de manera independiente. Para describir el pixel se puede utilizar cualquier característica relevante. Se puede modelar el valor de luminancia (definido como Y), el valor cromático (definido como RGB puesto que es la salida estándar de la mayoría de las cámaras, pero se puede utilizar cualquier otro espacio de color), la textura o variabilidad espacial del entorno del píxel, etc.

Para realizar una segmentación automática, se intenta crear un modelo de fondo robusto que modele la evolución del pixel a lo largo de todo el vídeo. Las técnicas de sustracción de fondo se basan en estos modelos y marcan como frente todo lo que no se adapta a él. Por tanto el problema queda reducido a crear un modelo que se adapte a la evolución temporal de cualquier característica que se esté modelando. Lamentablemente, esto no es trivial en la mayoría de los casos.

3.3 **Problemas del modelado de fondo.**

Uno de los principales problemas que se encuentran a la hora de modelar el fondo es la necesidad de crear el modelo sólo con las muestras disponibles hasta un determinado momento y suponiendo que las muestras futuras se parecen a las anteriores. Además se tiene que tener en cuenta la incapacidad para almacenar todas las muestras que han entrado al sistema, teniendo que utilizar un modelo que describa las muestras observadas y se vaya adaptando a las nuevas muestras. Finalmente, los algoritmos implementados tienen que ser eficientes computacionalmente ya que en multitud de aplicaciones se necesita que la actualización del modelo de fondo y la posterior segmentación de frente se realice en tiempo real. Teniendo en cuenta todas estas limitaciones, y según se describe en [10], un buen modelo de fondo debe solventar los siguientes problemas:

- *Cambios bruscos de iluminación*: son variaciones bruscas de iluminación, producidas por las distintas fuentes de luz presentes en la escena. Estas variaciones pueden ser erróneamente consideradas objetos pertenecientes al frente de la escena causando falsos positivos.

- *Sombras*: se producen por la interacción de las fuentes luminosas con los objetos que se encuentran en la escena. Generalmente se confunden como parte del frente y se suelen eliminar con técnicas de post-procesado.

- *Ruido*: el ruido introducido en el proceso de captación de las imágenes por las cámaras de video puede producir errores en la segmentación de objetos de video y es necesario eliminar su influencia en el resultado final.

- *Camuflaje*: este efecto aparece cuando los objetos del primer plano (del frente) poseen el mismo color y textura que el fondo (o la misma configuración de valores en

cualquiera de las características seleccionadas para el modelado); por este motivo, el frente se confunde o camufla como fondo.

3.4 La segmentación como un problema de clasificación.

El objetivo final de toda segmentación de vídeo es poder discriminar con una probabilidad de error pequeña, los píxeles de fondo de los píxeles de frente en cada cuadro del vídeo a analizar.

Este problema se puede ver como un problema de clasificación si se intenta hacer una distinción de los píxeles de fondo en distintas clases de pixel. Un ejemplo puede ser clasificar los píxeles de fondo entre fondo unimodal (estático) y fondo multimodal (dinámico), diferenciando ambos de los de frente. Esta clasificación se ha utilizado en este proyecto fin de carrera. Dicha distinción puede ser útil para adaptar el modelo de fondo de forma distinta o para realizar un tratamiento distinto a la hora de separar cada una de estas clases de pixel de los píxeles de frente.

Otro ejemplo de este tipo de esquemas puede consultarse en [11], donde se procesa cada pixel de manera diferente dependiendo de la clase a la que se ha asignado. Las posibles clases contempladas en ese trabajo son: frente, sombra, fantasma o ruido.

Para que las definiciones de fondo y frente realizadas en la sección 3.1 se entiendan mejor, a continuación se muestra la evolución temporal del valor de luminancia Y de una serie de píxeles. Al principio se muestra la evolución temporal de píxeles que siempre son fondo pero luego se muestra otra serie de píxeles que alternan frente y fondo. En estos últimos se puede observar algunos de los problemas (ver sección 3.3) que hacen que la segmentación no sea un proceso sencillo.

Pixel de fondo unimodal. Se ha analizado el pixel *(x=20,y=20)* del video *Silla.avi.* Para su localización espacial, el píxel aparece marcado con una cruz azul:

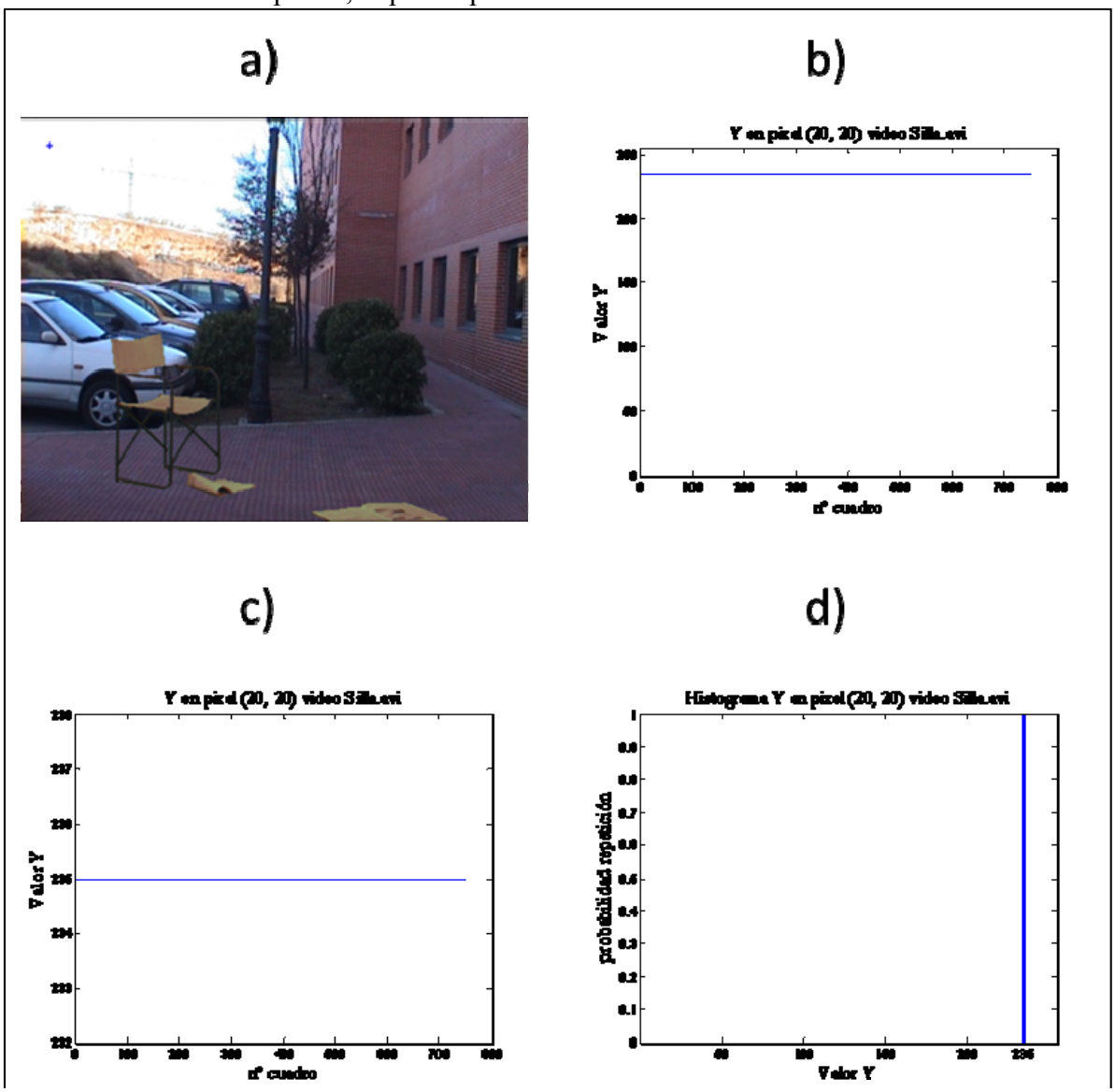


Figura 3-1: Evolución de *Y* en un pixel de fondo unimodal.

En a) se muestra el pixel estudiado. En b) se muestra *Y* en todos los cuadros del vídeo. En c) se dibuja nuevamente *Y* en un rango de valores acotado. En d) el histograma normalizado de *Y*.

Se observa que en este pixel el valor de luminancia es constante durante todo el vídeo. Por ello el modelado de este pixel es fácil y podría hacerse por medio de una Gaussiana Simple de media 235 y desviación típica igual a ±1. Si se quiere ser todavía más restrictivo se podría modelar por medio de un valor de luminancia constante igual a 235 y clasificar como fondo cuando el valor sea igual a 235 y como frente cuando el valor de

luminancia sea distinto a 235 (modelado mediante delta de Dirac discreta desplazada al valor 235).

Pixel de fondo multimodal. El modelado de un pixel de fondo no siempre es tan sencillo como en el caso anterior. En la siguiente figura se muestra la evolución temporal de *Y* de otro pixel. Aunque es fondo durante todo el video, la luminancia tiene mucha variación y no se puede modelar por medio de una Gaussiana Simple. Se ha analizado el pixel *(x=200,y=20)* del video *Jardín.avi* marcado con una cruz azul:

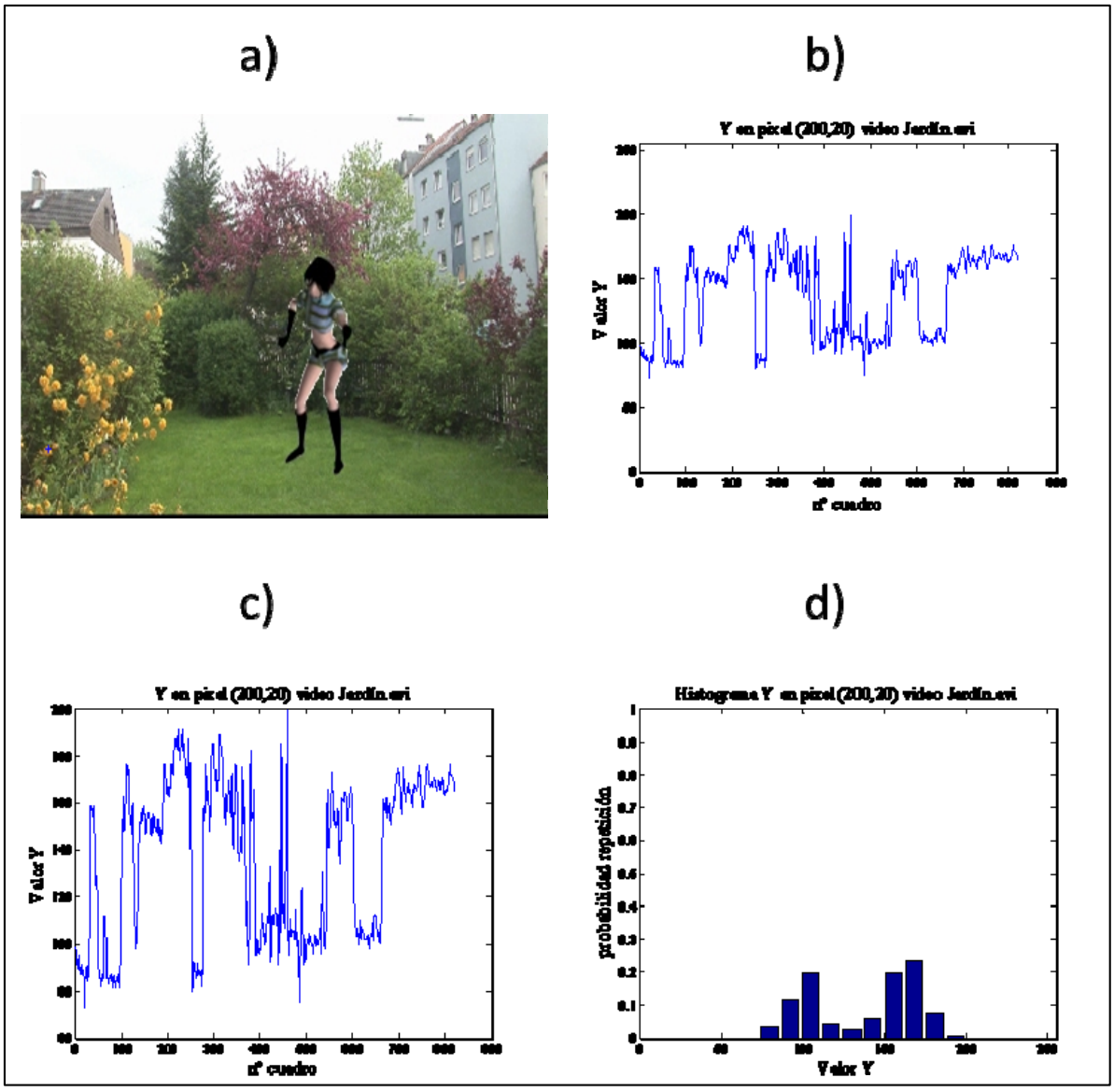


Figura 3-2: Evolución de *Y* en un pixel de fondo multimodal.

En a) se muestra el pixel estudiado. En b) se muestra *Y* en todos los cuadros del vídeo. En c) se dibuja nuevamente *Y* en un rango de valores acotado. En d) el histograma normalizado de *Y*.

Este pixel no se puede modelar por medio de un único modo ya que presenta apariencias que van desde el verde de la hoja del arbusto hasta el amarillo de la flor que se mueve delante. Esto es debido a que es un fondo multimodal producido por un arbusto agitado por el viento. Una buena forma de modelarlo es por medio de una Mezcla de Gaussianas como la propuesta en el algoritmo desarrollado por Stauffer y Grimson en [12].

Pixel de fondo unimodal por el que pasan objetos de frente. Se ha analizado el pixel *(x=220,y=220)* del video *Silla.avi* marcado con una cruz azul:

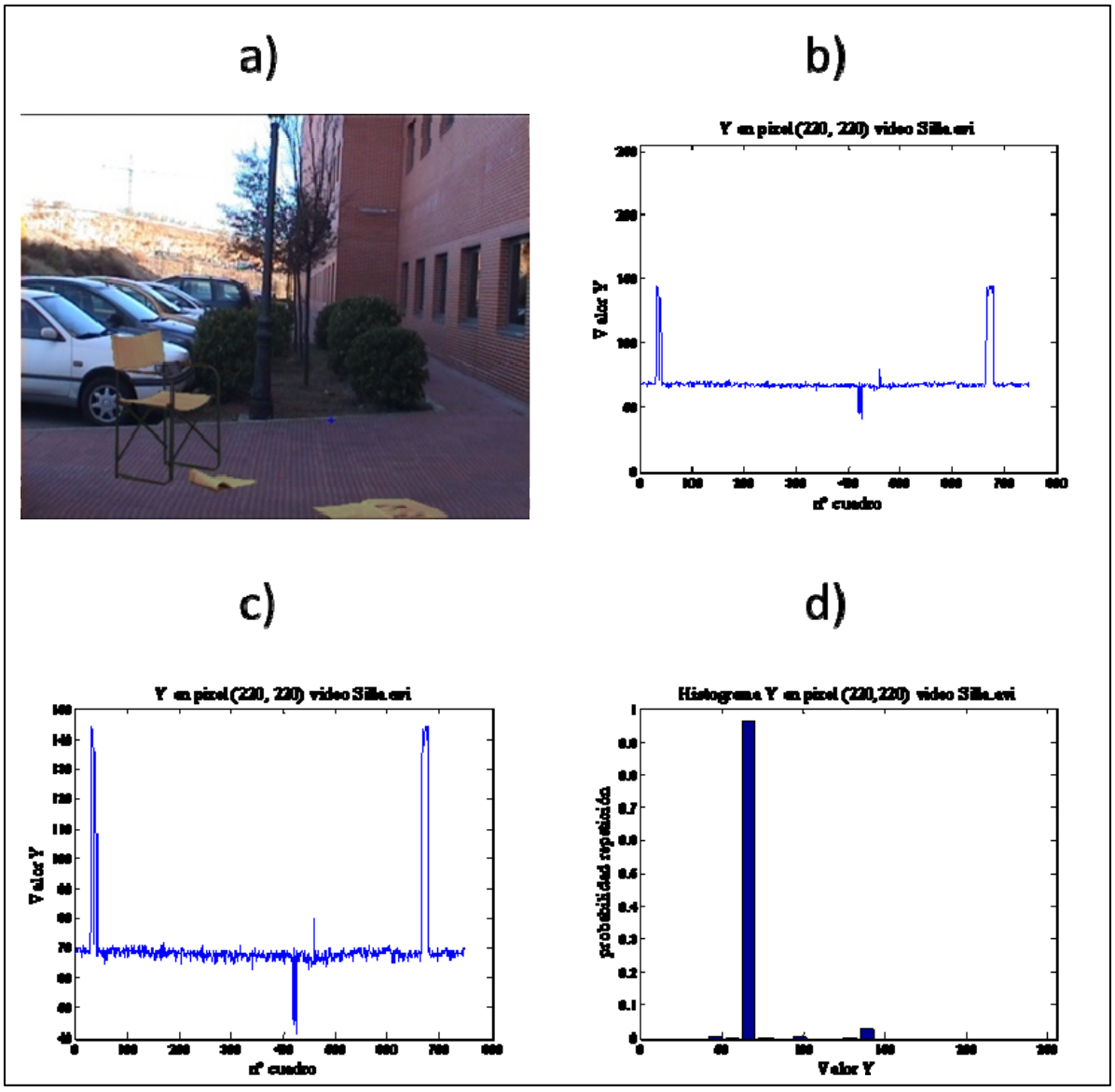


Figura 3-3: Evolución de *Y* en un pixel de fondo unimodal con presencia de frente.

En a) se muestra el pixel estudiado. En b) se muestra *Y* en todos los cuadros del vídeo. En c) se dibuja nuevamente *Y* en un rango de valores acotado. En d) el histograma normalizado de *Y*.

Este pixel se puede modelar por medio de una Gaussiana Simple con valor medio igual a 68 y una desviación típica igual a ±6. En la figura podemos observar que existen tres subidas del valor de luminancia por encima de 80 y una bajada por debajo de 50 que se corresponden con el paso de objetos de frente por este punto. Este tipo de pixel es fácil de modelar porque todo el fondo se describe mediante un único modo y los objetos de frente que pasan por él resultan en distancias grandes fáciles de discriminar.

Pixel de fondo multimodal por el que pasan objetos de frente. Se ha analizado el pixel *(x=200,y=20)* del video *Ingravidez.avi*, que aparece marcado con una cruz azul:

Se han utilizado aspas de color rojo para marcar las muestras que pertenecen al frente. En el histograma se han marcado las muestras de frente en rojo y las de fondo en azul para facilitar la visualización.

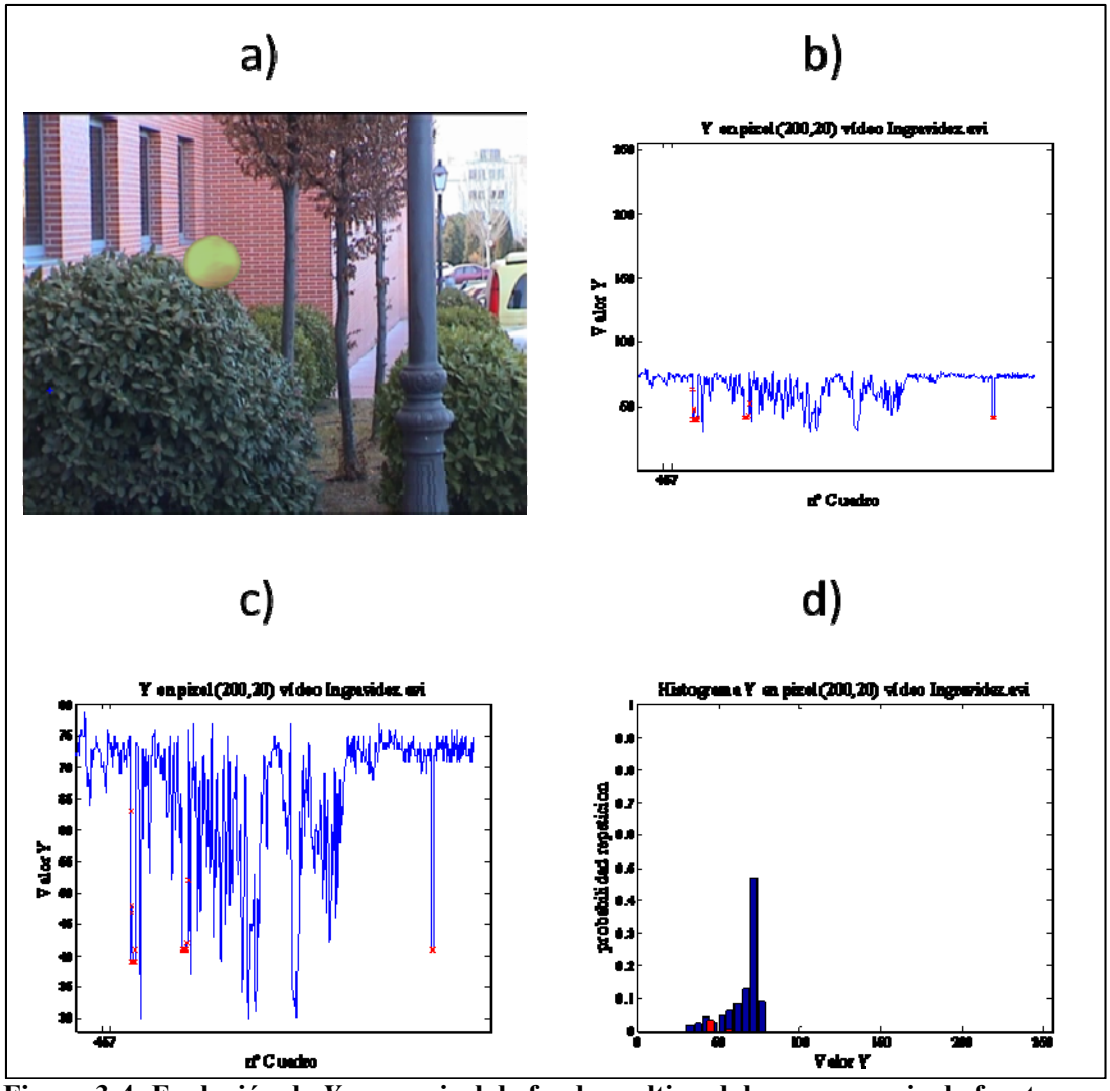


Figura 3-4: Evolución de *Y* en un pixel de fondo multimodal con presencia de frente.

En a) se muestra el pixel estudiado. En b) se muestra *Y* en todos los cuadros del vídeo. En c) se dibuja nuevamente *Y* en un rango de valores acotado. En d) el histograma normalizado de *Y*.

En este ejemplo se ve de manera clara alguno de los problemas que nos encontramos a la hora de realizar una clasificación de pixel a bajo nivel. En esta figura se puede observar como las muestras de frente resultan en distancias comparables a las del fondo. En el histograma se observa que las muestras de frente (rojo) se solapan con las de fondo (azul). Esto es debido a que existe camuflaje ya que hay cuadros que contienen un pixel de frente con la misma luminancia que otro cuadro que contiene un pixel de fondo. Por esta circunstancia algunos autores utilizan características adicionales a la hora de segmentar. Por ejemplo J.Gallego en [13] introduce un modelo de frente que ayuda a solventar problemas de camuflaje y oclusión de los objetos de frente. Kentaro en [10] expone los principales problemas que existen a la hora de modelar el fondo.

Para completar la lista de problemas, (ver sección 3.3), aparte del camuflaje y el ruido de cámara que se ponen en evidencia en la evolución del pixel ilustrada en la Figura 3-4, vamos a ver otras dos evoluciones donde se va a observar cambios de iluminación y existencia de sombras.

Pixel de fondo unimodal sometido a cambios locales de iluminación (sombras producidas por el frente). Se ha analizado el pixel *(x=30,y=30)* del video *highway.avi* marcado con una cruz azul:

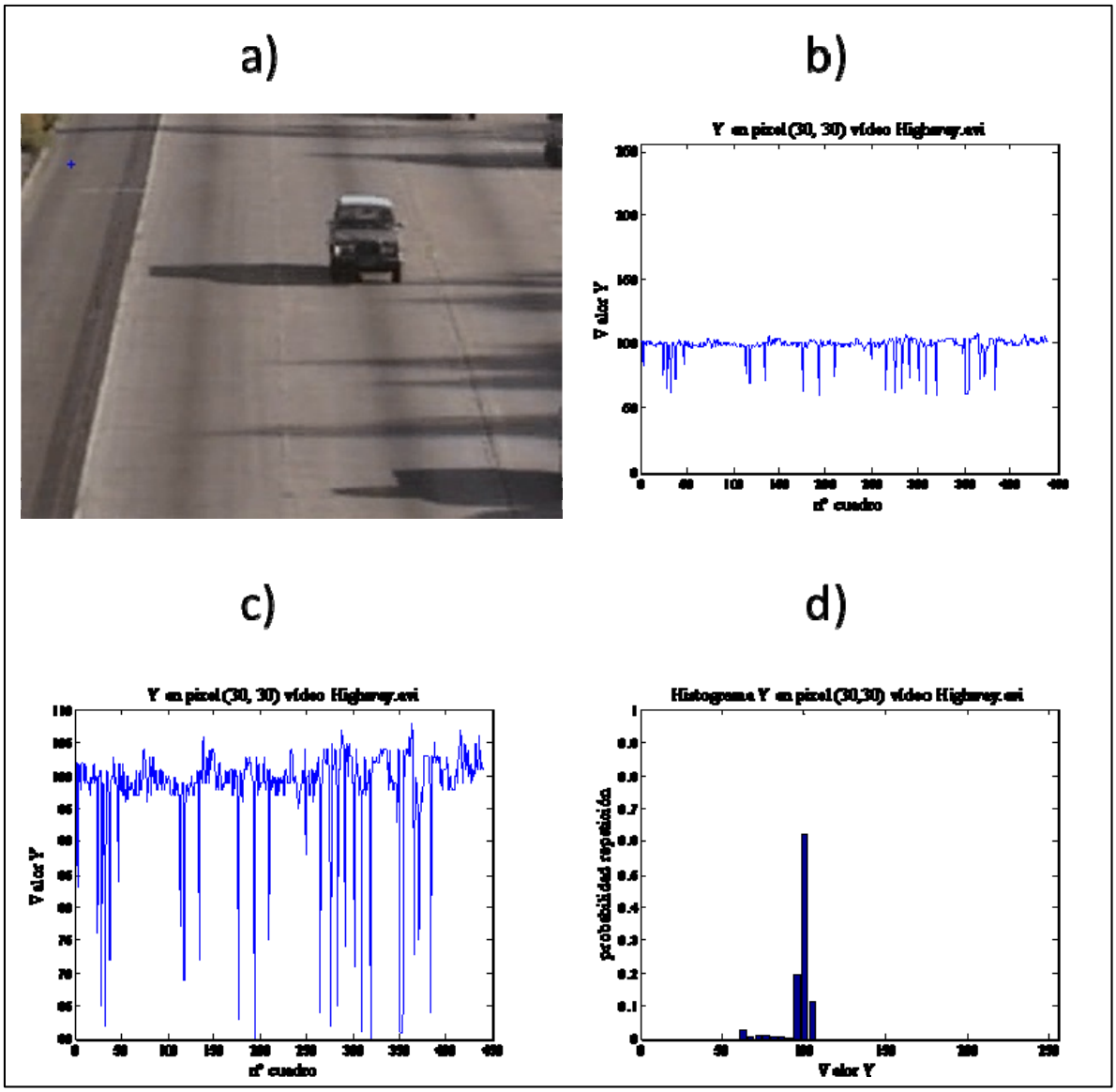


Figura 3-5: Evolución de *Y* en pixel de fondo unimodal con presencia de sombras.

En a) se muestra el pixel estudiado. En b) se muestra *Y* en todos los cuadros del vídeo. En c) se dibuja nuevamente *Y* en un rango de valores acotado. En d) el histograma normalizado de *Y*.

En este ejemplo se ve que existe un modo principal para valores de *Y* entre 95 y 105 y que luego aparecen otros modos secundarios a luminancias más bajas que se corresponden con las sombras del vídeo. En [14] se intenta solucionar este problema utilizando texturas.

Para terminar vamos a analizar otro de los problemas fundamentales que suelen aparecer en muchos vídeos a la hora de segmentar. Se trata de los cambios bruscos de iluminación. La mayor parte de los algoritmos son capaces de adaptarse a cambios graduales pero los cambios bruscos se comportan de forma parecida al frente que pasa por la escena generando nuevamente un problema en la detección de la clase a la que pertenecen los píxeles afectados.

Pixel de fondo unimodal con cambio brusco de iluminación. Se ha analizado el pixel *(x=20,y=20)* del video *Lobby.avi* marcado con una cruz azul:

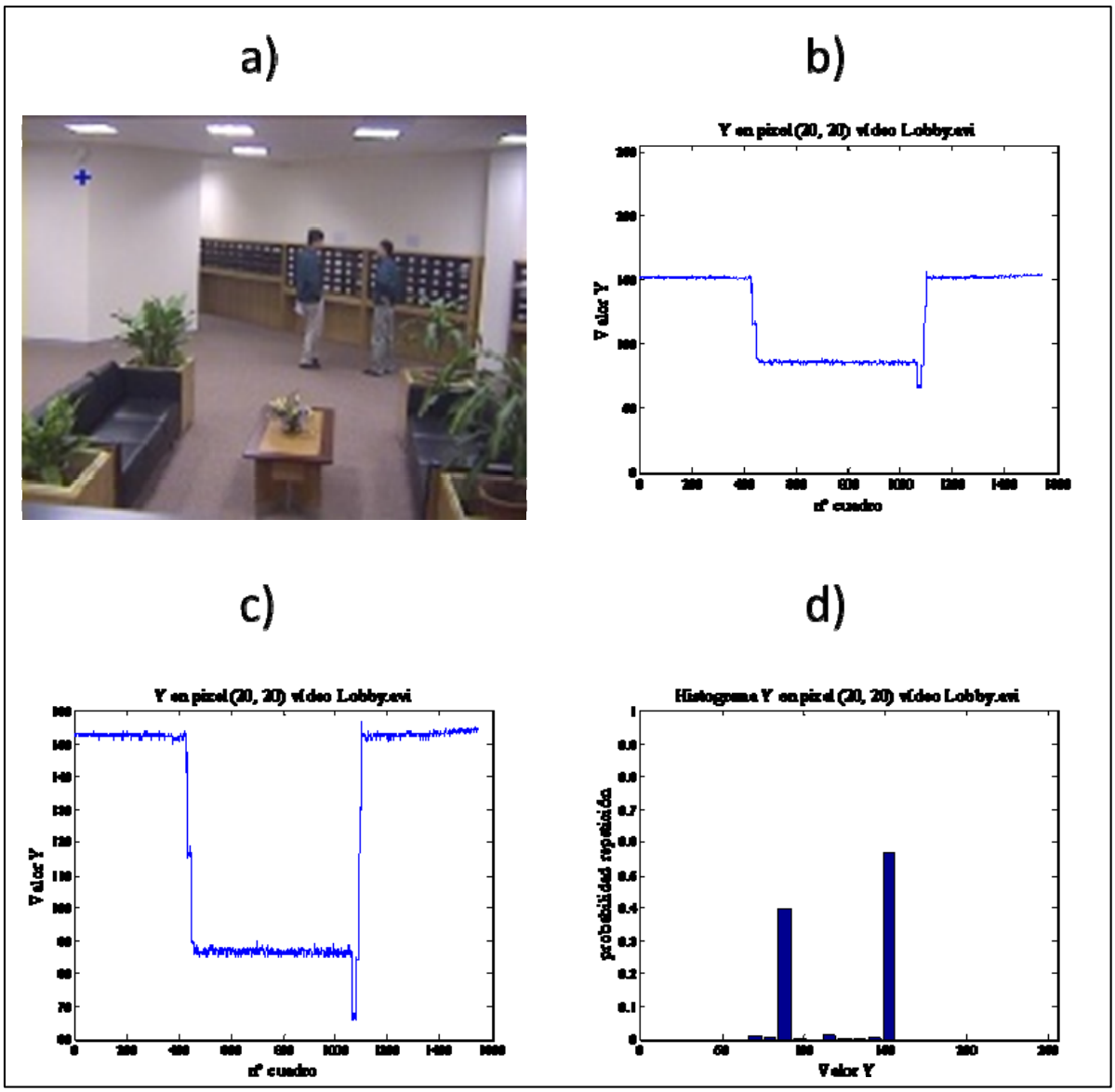


Figura 3-6: Evolución de *Y* en pixel fondo unimodal con cambios bruscos de iluminación.

En a) se muestra el pixel estudiado. En b) se muestra *Y* en todos los cuadros del vídeo. En c) se dibuja nuevamente *Y* en un rango de valores acotado. En d) el histograma normalizado de *Y*.

El problema del cambio brusco de iluminación es que hay que buscar un algoritmo que sea capaz de adaptarse a él, pero que, al mismo tiempo, sea capaz de discriminar cuando el cambio se ha producido por la aparición de frente.

3.5 Modelos de sustracción del fondo.

Las técnicas de segmentación basadas en sustracción de fondo, inicializan un modelo de fondo lo más completo posible que pueda ajustarse a las características del vídeo analizado y que clasifique como frente los pixeles que no se ajustan a este modelo. Todos estos pixeles forman la máscara de frente que suele ser el resultado final de la segmentación. Podemos ver algunos ejemplos en [2], [3], [15] y [16].

Aquellos píxeles no clasificados como frente, es decir los que se ajustan al modelo se clasifican como fondo y sirven para actualizar y mantener el modelo de fondo.

3.5.1 Esquema general de sustracción del fondo. (BS)

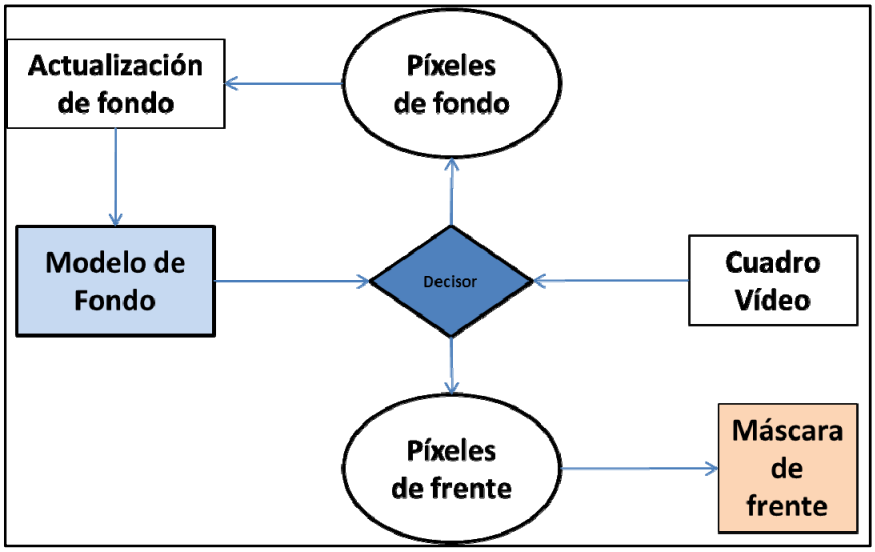


Figura 3-7: Esquema de sustracción de fondo (*BS*).

3.5.2 **Modelado de fondo.**

La forma de definir el fondo es la tarea más importante dentro de un algoritmo de segmentación por substracción de fondo, ya que la segmentación se realiza mediante comparaciones contra este modelo.

En el modelado de fondo se intenta recoger la evolución de la escena durante toda la duración del vídeo. Se debe tener en cuenta que podemos tener problemas a la hora de realizar el modelado como los descritos en la sección 3.3.

Atendiendo a los criterios propuestos por Cristani [17], podemos separar el modelado del fondo en modelo de inicialización y modelo de adaptación:

Modelo de inicialización.

Existen varias formas de afrontar este problema. Una de ellas es tomar como fondo la primera imagen del vídeo [1]. Puede dar problemas ya que puede contener frente (a este problema se le denomina *inicio en caliente*). Además, este método de inicialización sería poco flexible, puesto que puede presentar apariencias del fondo ocluidas, lo que ralentizaría la obtención de un modelo útil.

Otra opción es tomar varias imágenes del vídeo (n imágenes) y reestimar su apariencia más común mediante el cálculo de su media o su mediana con el fin de utilizar la imagen resultante cómo modelo de fondo [18].

Finalmente pueden utilizarse dos cuadros separados en el tiempo (p.e.1 segundo) umbralizando la diferencia entre ellos hasta que un porcentaje de los píxeles no supere un umbral (impuesto por dicho porcentaje). Actualizando los valores de este fondo en construcción, y repitiendo este proceso varias veces, el modelo de fondo construido será cada vez más robusto [19].

Modelo de adaptación.

Dentro del modelado de fondo siempre hay que tener en cuenta su actualización, ya que el fondo no suele permanecer constante durante el video, y como consecuencia, el modelo inicial puede no ser válido o útil para la segmentación. En la sección 3.3 se citan

problemas y en la sección 3.4 ejemplos de evoluciones de píxeles de fondo que evidencian la necesidad de actualizar el modelo de fondo.

Existen múltiples técnicas para adaptar el modelo de fondo: el método más sencillo es ir actualizando el modelo de fondo directamente con la información disponible en el **cuadro anterior** [20]. Esta técnica es muy sencilla pero da problemas ya que se introducen objetos de frente en el modelo de fondo. Es decir no se puede modelar el fondo que está ocluido por un objeto de frente.

Otra técnica que soluciona el problema anterior sería actualizar cada instante el modelo de fondo a través de la **media** o la **mediana** de los valores de las imágenes anteriores a dicho instante. Existen técnicas de mayor coste computacional que dan mejores resultados como las que mantienen un buffer para cada píxel con los valores que dicho píxel ha tomado en instantes anteriores y actualizan el modelo de fondo en función de los valores de este buffer (por ejemplo, estimando funciones de probabilidad con los valores almacenados en el buffer). Como ejemplo de este modelo de actualización podríamos citar el trabajo realizado por Koller en el artículo [21], que utiliza un filtro basado en una ventana de promedio temporal que permite detectar los objetos en movimiento cuando se producen variaciones de iluminación.

El método más conocido en el estado del arte es el de **media móvil** (*Running average*). Su amplia utilización radica en el hecho de que consigue adaptarse a los cambios progresivos que se producen manteniendo además un bajo coste computacional. Mediante esta operación, el modelo de fondo se adapta a las variaciones progresivas que puedan producirse en la escena, como por ejemplo, variaciones en la iluminación solar a lo largo del día en entornos exteriores. Para un pixel, la actualización del fondo se realiza como una media ponderada entre la media almacenada en el modelo de fondo (m_t) y la muestra actual (x_t):

$$m_t = \alpha x_t + (1-\alpha) m_{t-1} \qquad \textit{eq 3.1}$$

La velocidad de adaptación viene modelada por el parámetro α. En modelos conservadores α tiende a ser pequeña ya que se da poco peso a las nuevas muestras. En

modelos donde se busca una adaptación más rápida a los cambios, α es mayor para dar más peso a las nuevas muestras. Un ejemplo de este tipo de técnica se propone en [22].

Recientemente, se han utilizado métodos bayesianos ([23], [24] y [25]) para el proceso de adaptación del modelo de fondo. En estos casos, los modelos siguen un aprendizaje evolutivo siguiendo esquemas de aprendizaje fundamentados en la regla de Bayes.

3.5.3 **Técnicas concretas de modelado no Bayesianas.**

En esta sección, clasificaremos las distintas técnicas en función del tipo de fondo que intenten modelar 3.1. Podemos clasificar los métodos de segmentación en métodos que utilizan un modo (fondo unimodal) y métodos que utilizan dos o más modos (fondo multimodal) para definir el fondo.

3.5.3.1 **Un modo. Fondo unimodal.**

3.5.3.1.1 **Objetivos.**

Estos métodos intentan modelar el fondo incluyendo toda la información disponible en un único modo. Se pueden adaptar a cambios progresivos pero están limitados ya que no soportan cambios bruscos entre imágenes consecutivas.

El objetivo principal de estas técnicas de modelado de fondo es obtener una imagen lo más parecida posible al verdadero fondo de la escena para comparar contra ella la imagen de entrada del vídeo y poder diferenciar entre fondo y frente.

3.5.3.1.2 **Ejemplos.**

Gaussiana simple.

El método de la Gaussiana Simple (*SG*) representa cada píxel x_t con una distribución unimodal Gaussiana definida por dos parámetros: media μ_t y varianza σ_t^2. La varianza permite modelar los pequeños cambios que ocurren en la imagen.

$$\left.\begin{aligned} \mu_t &= \sum_{i=1}^{i=T} \frac{x_i}{t} \\ \sigma_t^2 &= \sum_{i=1}^{i=T} \frac{x_i^2}{t} - \mu_t^2 \end{aligned}\right\} \forall \quad x_t \in BG_t \qquad \textit{eq 3.2}$$

En cada instante t, se determina si un píxel pertenece al modelo de fondo BG_t si el valor de dicho píxel en la imagen cae dentro de la Gaussiana definida para ese píxel, es decir, si la diferencia entre el valor del píxel y el de la media modelada para dicho píxel μ_t es inferior a *K* veces la desviación típica σ_t .

Una gran cantidad de algoritmos utilizan como base de su segmentador de objetos en movimiento el principio de la Gaussiana Simple. Es el caso de Gordon [26], que genera para cada canal de color (*RGB*) de un píxel una distribución Gaussiana de matriz de covarianza diagonal.

El método de la *SG* no es capaz de adaptarse a fondos multimodales, en los que cada píxel de fondo puede tomar valores muy diferentes, sin por ello, dejar de ser un píxel de fondo.

3.5.3.2 **Dos o más modos. Fondo multimodal.**

3.5.3.2.1 **Objetivos.**

Estos métodos intentan crear un modelo de fondo más rico donde cada píxel de fondo puede estar caracterizado con más de un modo.

Son muy útiles para modelar fondos multimodales. Un mismo píxel puede contener dos modos muy distintos. Por ejemplo en uno se modela una hoja verde y en otro el cielo azul que se encuentra detrás de la hoja y se descubre cuando esta se mueve. Esto puede ser necesario si a lo largo del vídeo se ve la hoja o el cielo dependiendo del momento y del viento de la escena. En estos métodos no se puede caracterizar cada píxel del fondo de la escena con un único modo.

3.5.3.2.2 **Ejemplos**

Mezcla de Gaussianas.

En fondos multimodales, que contienen objetos no estáticos, tales como hojas de árboles en movimiento, olas, etc., hay píxeles cuyos valores de intensidad varían entorno a un conjunto finito de valores característicos. Por este motivo, un píxel no puede modelarse por medio de un valor (una media) y un conjunto en torno a éste (la desviación típica) utilizando una distribución Gaussiana. La Mezcla de Gaussianas (*MoG*) propone una solución a este problema [27] que consiste en modelar la intensidad de los píxeles con una mezcla de *k* distribuciones Gaussianas (donde *k* es un número pequeño, frecuentemente se utiliza de 3 a 5 dependiendo de la multimodalidad del escenario de análisis) definidas por los siguientes parámetros: media $\mu_{k,t}$, varianza $\sigma^2_{k,t}$, y peso $w_{k,t}$. La función de distribución de la probabilidad del modelo queda:

$$P(BG_t) = \sum_{j=1}^{j=k} w_{j,t} N(\mu_{j,t}, \sigma^2_{j,t}) \qquad \textit{eq 3.3}$$

En *MoG* se evalúa la pertenencia de los nuevos píxeles de la imagen a todo el modelo. Si se encuentra pertenencia, se actualizan los parámetros del modelo para ese píxel: la media $\mu_{k,t}$, la varianza $\sigma^2_{k,t}$ y el peso $w_{k,t}$ y se caracteriza el pixel como fondo. Si no se parece a ninguna de las distribuciones asociadas al píxel, la de menor peso se sustituye por una nueva Gaussiana de media el valor del píxel en la imagen actual y varianza un valor pequeño.

En el modelo de Mezcla de Gaussianas, la combinación de k distribuciones Gaussianas cuya probabilidad de ocurrencia (suma de pesos $w_{k,t}$) supere un determinado umbral, denominado umbral de frente, permitirá modelar cada píxel del fondo en cada instante.

Al igual que sucede en el modelo de la *SG*, se han desarrollado muchas variantes al esquema clásico de *MoG* definido en [27]. Por ejemplo, Javed [16] utiliza un modelo de *MoG* para llevar a cabo la substracción de fondo en cada canal de color, relacionando dichos canales a través de una matriz de covarianza. Harville [28], propone un método para

modelar el fondo que combina la *MoG* con la información de profundidad y luminancia; este método se utiliza en detección de automóviles.

No obstante, la *MoG* también posee inconvenientes. En primer lugar, conlleva una alta carga computacional. Por otro lado, es muy poco robusta a cambios repentinos de iluminación. Además, algunos fondos multimodales requieren un número de distribuciones k elevado para modelar cada píxel, lo cual implica un incremento en la carga computacional. En este método de representación del fondo es muy importante la forma de actualizar las medias y las varianzas para adaptarse a los cambios del fondo.

A pesar de estos inconvenientes, la *MoG* es capaz de manejar una distribución multimodal de fondo ya que mantiene una función de densidad de probabilidad para cada píxel. Al ser un método paramétrico, puede describir apariencias complejas del fondo sin necesidad de actualizar un gran buffer de almacenamiento de imágenes como requieren los métodos no paramétricos tales como el *KDE* [29].

3.5.4 **Modelado del fondo por esquema Bayesianos multicapa.**

Nos centraremos en los modelos Bayesianos multicapas, dado que son los utilizados en el proyecto final de carrera que se presenta. Estos métodos de actualización de fondo mediante esquemas Bayesianos, inicializan generalmente un número suficiente de capas donde se modela independientemente cada una de las apariencias del píxel que se observan en el vídeo. Las capas se van modelando según van entrando nuevos píxeles al sistema. Cada nuevo pixel se va comparando con cada una de las capas inicializadas comenzando siempre por la más probable. Cuando se encuentra pertenencia a una de ellas el proceso comparativo finaliza y se actualiza con el nuevo valor.

Estos algoritmos utilizan la regla de Bayes para modelar las probabilidades no conocidas, sabiendo o estimando las probabilidades condicionadas y las probabilidades a priori. La regla de Bayes, puede expresarse como:

$$P(A_i \mid B) = \frac{P(B \mid A_i)P(A_i)}{\sum_{j=1}^{n} P(B \mid A_j)P(A_j)} \qquad \textit{eq 3.4}$$

Donde $P(A_t)$ representan las probabilidades a priori, $P(B \mid A_t)$ es la probabilidad de B condicionada a la hipótesis A_t y $P(A_t \mid B)$ son las probabilidades a posteriori de cada hipótesis.

El objetivo final de un modelado de fondo (*BG*) o de frente (*FG*) y de su posterior adaptación es encontrar un modelo de la muestra x_t y la probabilidad asociada de que sea parte del frente (*PoFG*) o del fondo (*PoBG*).

Comparemos el proceso con la formulación Bayesiana común a los algoritmos que modelan el fondo mediante una sola distribución de probabilidad conjunta (como puede ser una *MoG*) y no mantienen ningún modelo del frente.

Dada la muestra x_t, la probabilidad de que esa muestra sea fondo puede escribirse mediante la aplicación de la regla de Bayes simple como:

$$P(BG_t \mid x_t) = \frac{P(x_t \mid BG_t)P(BG_t)}{P(x_t)} \qquad \textit{eq 3.5}$$

, mientras que la probabilidad de que sea frente quedaría:

$$P(FG_t \mid x_t) = \frac{P(x_t \mid FG_t)P(FG_t)}{P(x_t)} \qquad \textit{eq 3.6}$$

Dividiendo (para eliminar la $P(x_t)$), resultaría:

$$\frac{P(BG_t \mid x_t)}{P(FG_t \mid x_t)} = \frac{P(x_t \mid BG_t)P(BG_t)}{P(x_t \mid FG_t)P(FG_t)} \qquad \textit{eq 3.7}$$

O, lo que es lo mismo:

$$x_t \begin{cases} \in BG & si \quad \dfrac{P(x_t \mid BG_t)P(BG_t)}{P(x_t \mid FG_t)P(FG_t)} \geq 1 \\ \in FG & si \quad \dfrac{P(x_t \mid BG_t)P(BG_t)}{P(x_t \mid FG_t)P(FG_t)} < 1 \end{cases} \qquad \textit{eq 3.8}$$

Donde, priorizamos la asignación al fondo. Como además, el frente no suele ser modelado, puede asumirse que presenta una distribución uniforme entre todos los valores del rango de x_t, pongamos $P(x_t \mid FG_t)P(FG_t) = \theta \ \forall x_t$, así, finalmente quedaría:

$$x_t \in BG \quad si \quad P(x_t \mid BG_t)P(BG_t) \geq \theta \qquad \textit{eq 3.9}$$

Donde, exportando a un modelo clásico de *MoG*, la distribución de densidad de probabilidad de las Gaussianas estaría relacionada con la evaluación de valores particulares para la imagen de entrada mediante $P(x_t \mid BG_t)$, mientras que el conjunto de los pesos de la Gaussiana se relacionaría con $P(BG_t)$.

Otra alternativa sería discriminar entre frente y fondo en función del valor de $P(BG_t)$ (relacionado con el peso de la Gaussiana en un modelo *MoG*). Sólo aquellas Gaussianas bien entrenadas (es decir con muchas muestras observadas) modelarían el fondo, mientras que el resto, se asociarían al frente.

Como puede observarse, en ambas alternativas, $P(BG_t)$ puede representarse, (y de hecho es representada) mediante una función de densidad de probabilidad (*fdp*), con la condición inherente de que la suma de los valores de esta *fdp* en todo su rango han de sumar 1. Por lo tanto, alteraciones de la probabilidad de alguno de sus valores (pongamos por ejemplo que se incrementa el peso de una de las Gaussianas de la mezcla) alteran el resto de los valores en el rango (o los pesos de las otras Gaussianas).

Esta situación deriva en la necesidad de plantear esquemas de actualización que sean capaces de adaptarse rápidamente a apariencias marginales que adquieren rápidamente más presencia, al mismo tiempo que mantienen adecuadamente las apariencias predominantes no observadas durante un tiempo.

Si, por el contrario definimos un modelo de fondo multicapa, podemos introducir una variable discreta en la formulación: z_t. Esta variable puede tomar los valores $z_t \in \{1,.....,k\}$. Esta configuración busca representar el modelo de fondo (BG_t) en k clases o capas. Con el objetivo de estimar la probabilidad de que, dada la observación x_t, los valores de BG_t se modelen en cada una de las k capas: (BG_t, z_t) podemos utilizar la siguiente aproximación bayesiana propuesta en [30]:

$$p(BG_t, z_t \mid x_t) = p(BG_t \mid z_t, x_t)\, p(z_t \mid x_t) \qquad \textit{eq 3.10}$$

De esta manera, podemos aislar el modelado interno a la capa (o intra-capa) $p(z_t \mid x_t)$ (equivalente a la $p(BG_t \mid x_t)$ descrita en la ecuación *eq 3.5*) del modelado del fondo entre capas (o inter-capa) $p(BG_t \mid z_t, x_t)$, que indica la probabilidad, de que dada una capa a la que pertenece la muestra de entrada, ésta pertenezca al modelo de fondo.

Esta división permite una mayor flexibilidad en dos procesos clave:

- En el modelado de los procesos de actualización y discriminación de cada capa (la actualización de una no requiere necesariamente de la actualización de todas).

- En el proceso de asignación de fiabilidad a cada capa.

Tras calcular la probabilidad de pertenencia a cada capa, se asignará en función de la descripción de la capa en el modelo cuáles de estas capas pertenecen o no al fondo (pertenencia ya calculada), situación que depende tanto de la capa, como de las apariencias observadas hasta ese momento.

En el modelo interno a la capa se podrían utilizar diferentes técnicas de modelado tanto paramétricas como *SG*, *MoG*, como no paramétricas, por ejemplo *KDE* [29] para determinar la verosimilitud de las nuevas muestras a la capa $p(z_t \mid x_t)$.

En este proyecto final de carrera, se modelará un único modo en cada capa, por lo que la técnica elegida será el uso de distancia de Mahalanobis para la evaluación del proceso de pertenecía, y el uso de un modelado basado en *SG* para la umbralización de la distancia (tal y como se detallará en el capítulo 4).

La solución elegida para el cálculo de la probabilidad de que la muestra pertenezca a una de las *k* capas definidas en z_t, $p(z_t \mid x_t)$ se define en el capítulo 4 y se proponen mejoras sobre ésta tanto en el cálculo de pertenencia a la capa (sección 6.2) como en los procesos de actualización e inicialización de los parámetros del modelado (sección 7.2).

Por otro lado, en el capítulo 5 se describe un diseño basado en el estado del arte para el proceso de modelado entre capas, estimación de la $p(BG_t \mid z_t, x_t)$, (o simplemente,

modelado de fondo). A este sistema lo denominaremos *Bayesiano Básico*, por ser el punto de partida sobre el que se realiza el resto del proyecto final de carrera.

Sobre este diseño se realizarán cambios en la definición de las clases de pixel (sección 6.3.1) y se introducirán nuevas características que permitirán esta clasificación.

3.5.4.1 **Ejemplos de técnicas de modelado Bayesiano Multicapa.**

Estos métodos se caracterizan por utilizar un modelo de adaptación basado en aprendizaje bayesiano descrito en la sección anterior 3.5.4 como en [22], [23] y [25].

Dependiendo del autor se pueden definir varias clases, cada una modelada en una capa diferente, y se estima la probabilidad de pertenencia de las nuevas muestras a cada una de las clases. Por ejemplo C.Benedek en [31] utiliza este método de adaptación y crea tres clases distintas: frente, fondo y sombra.

En [1] solo se modela el fondo y se crea una capa distinta dentro del modelo para cada una de las distintas apariencias que puede tomar un pixel (fondos multimodales definidos en 3.1). Hay que destacar que cada una de las capas se modela por medio de *SG* descrita en 3.5.3.1.2, que se va actualizando mediante aprendizaje bayesiano. En el artículo, se introduce una nueva medida que se ha utilizado durante todo el trabajo realizado en este proyecto. Se trata de la medida de confianza definida desde ahora como *C* y que nos indica como es de fiable cada capa almacenada en el modelo generado (Se destaca que en cada capa se modela la apariencia para un único pixel. Hay que entender el concepto de capa a nivel de pixel y no como una imagen completa).

Por el contrario, en [32] se crea un modelo de fondo mediante métodos no paramétricos para modelar fondos multimodales que compite contra un modelo de frente. En este trabajo se modela el fondo multimodal utilizando un esquema Bayesiano en una sola capa utilizando métodos no paramétricos al contrario que en [1] donde se modelan los fondos multimodales mediante esquema Bayesiano en diferentes capas utilizando en cada capa *SG*.

En [1] se modela la misma clase (fondo) en diferentes capas, en cambio en [32] se modelan distintas clases (frente y fondo) en diferentes capas.

3.5.4.2 **Comparativa Bayesiano vs Mezcla de Gaussianas.**

Como se ha mencionado en la sección 3.5.4, una de las ventajas que ofrece este método es que modela cada una de las capas de forma independiente al resto, al contrario que por ejemplo la *MoG* donde cuando se crea una nueva Gaussiana se baja el peso de las que ya existían previamente. El proceso de actualización de las capas de un modelo Bayesiano solo afecta o bien a la capa a la que se adapta la muestra o bien a la capa de menor relevancia que es eliminada para crear una nueva si la muestra no se adapta a ninguna de las capas existentes. Los pesos o relevancia de las otras capas no se alteran.

La probabilidad de pertenencia al fondo en un sistema Bayesiano multicapa se evalúa independientemente en cada capa previa ordenación de ellas. En un esquema *MoG* en cambio se calcula como el sumatorio de las probabilidades de pertenencia a cada Gaussiana ponderados por el peso de la Gaussiana.

Por todo ello el modelado Bayesiano multicapa ofrece la ventaja de modelar cada capa de manera independiente al resto consiguiendo con ello un modelado por modo que no afecta al resto de los modos que aparecen en escena como en [33].

3.5.5 **Discriminación frente fondo.**

Para poder decidir si un pixel pertenece al frente o al fondo hay que establecer unas reglas que permitan medir el grado de similitud o la probabilidad de pertenencia a una u otra clase. De esta manera es necesario por un lado diseñar un modelo contra el que comparar de entre los explicados en la sección anterior 3.5.3 u otros similares, y por otro tener una medida objetiva que indique el parecido al modelo. Para realizar esta comparación se puede utilizar distancia Euclídea (d), de Mahalanobis (D) u otras como la CityBlock [34] que no explicaremos en este proyecto final de carrera.

Estas distancias miden el parecido del valor modelado a la nueva muestra de entrada, de cualquier característica que se pueda modelar. Algunos ejemplos de características

utilizadas por el modelado son la luminancia Y [28], el color [1] (generalmente organizado siguiendo el esquema *RGB*), la textura [35] o la distribución del color etc.

La siguiente sección introduce brevemente las diferentes estrategias para medir similitudes entre características utilizadas durante el desarrollo de este proyecto.

3.5.5.1 **Similitud al modelo.**

La **distancia Euclídea** es la distancia entre dos puntos de un espacio Euclídeo. En ella se resta por separado cada una de las componentes que forman las muestras a comparar y se hace la raíz cuadrada de todas ellas (Teorema de Pitágoras). Estructuralmente un **espacio euclídeo** es un espacio vectorial normado sobre los números reales de dimensión finita. Se define por la siguiente fórmula:

$$d((x_1, y_1),(x_2, y_2)) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \qquad \textit{eq 3.11}$$

Una alternativa a esta distancia es aquella que contempla la distribución de los datos en espacios de características multidimensionales, considerando la correlación entre las características utilizadas para el modelado. Dado un pixel *(x,y)* en el instante *t* del cuadro de entrada se obtiene el vector $\vec{x}_t$ (vector *N* dimensional), **la distancia de Mahalanobis** *D* entre el pixel y su modelo con media $\vec{m}_t$ y covarianza Σ_t queda:

$$D_t^2 = (\vec{x}_t - \vec{m}_t)^T \Sigma_t^{-1} (\vec{x}_t - \vec{m}_t) \qquad \textit{eq 3.12}$$

La utilidad de la **distancia de Mahalanobis** radica en que es una forma de determinar la similitud entre dos variables aleatorias multidimensionales que, a diferencia de la distancia Euclidea, considera la correlación entre ellas. El uso de la distancia de Mahalanobis y el cálculo de matrices de covarianza se ha utilizado en trabajos como [36], [37] y [38].

Matriz de covarianza: Es la generalización natural a dimensiones superiores del concepto de varianza de una variable aleatoria *N* dimensional $\vec{X}$, siendo $\vec{M} = E[\vec{X}]$.

$$\sum = \begin{pmatrix} E\left[(X_1 - M_1)(X_1 - M_1)\right] & \cdots & E\left[(X_1 - M_1)(X_N - M_N)\right] \\ \vdots & \ddots & \vdots \\ E\left[(X_N - M_N)(X_1 - M_1)\right] & \cdots & E\left[(X_N - M_N)(X_N - M_N)\right] \end{pmatrix} \qquad \textit{eq 3.13}$$

El sistema basado en **mezcla de expertos** es otro esquema de discriminación en el cual se calcula por separado la distancia en cada una de las dimensiones. Posteriormente se decide por mayoría, con funciones lógicas basadas en operaciones de *OR* o *AND* o con soluciones más complejas como el uso de máquinas de soporte de vectores (*SVM*) o redes neuronales (*ANN*) entre los diferentes expertos. A continuación se muestra la fórmula para calcular la distancia entre dos puntos con *N* dimensiones. En este caso se utilizan *N* expertos.

$$d(\vec{x}_t - \vec{m}_t) = \begin{cases} x_{t,1} - m_{t,1} \\ x_{t,2} - m_{t,2} \\ \ldots\ldots\ldots \\ x_{t,N} - m_{t,N} \end{cases} \qquad \textit{eq 3.14}$$

No es el objetivo de este trabajo el ahondar en los diferentes esquemas de fusión de expertos. Dos ejemplos de uso de mezclas de expertos pueden consultarse en [39] y en [40].

3.6 **Tratamiento a nivel de blob.**

Para intentar eliminar problemas derivados del tratamiento a nivel de pixel en numerosos trabajos se introduce tratamiento a nivel de blob que permite introducir nueva información teniendo en cuenta la consistencia espacial entre píxeles vecinos solventando de esta manera clasificaciones erróneas del píxel. Los blobs son el resultado de realizar un estudio de conectividad en la máscara de frente. Cada blob se corresponde con una componente conexa.

Como se indica en [41] el introducir información a nivel de regiones permite reducir el número de falsos positivos y permite realizar una descripción más robusta de la escena. De esta manera aunque se presente un falso positivo en alguno de los píxeles del blob, puede solventarse el error gracias a los vecinos que lo rodean.

J.Carmona en [11] utiliza información a nivel de blob para clasificar los píxeles de entrada y realizar un tratamiento individualizado dependiendo del resultado de la clasificación.

El trabajar con este tipo de información puede ayudar a eliminar los problemas descritos en 3.3. Además se puede utilizar este tipo de información para tomar decisiones de manera conjunta para un grupo de píxeles conexos siendo de esta manera más robustos a los problemas derivados de trabajar con píxeles sueltos.

Una ventaja derivada de la utilización del blob como unidad de análisis es la utilizada en [1], donde se elimina ruido impulsivo mediante filtrado por tamaño. El proceso es el siguiente: se define un tamaño mínimo de blob para considerarlo como frente. Todas las componentes conexas que tengan un tamaño menor que el mínimo exigido se consideran como fondo y se eliminan de la máscara de frente. De esta manera el sistema separa las zonas que contienen frente y utiliza la información de toda la zona para tomar decisiones.

A continuación se muestra una figura en la que se muestra la extracción de blobs en el cuadro 194 del vídeo *BaileUni.avi*:

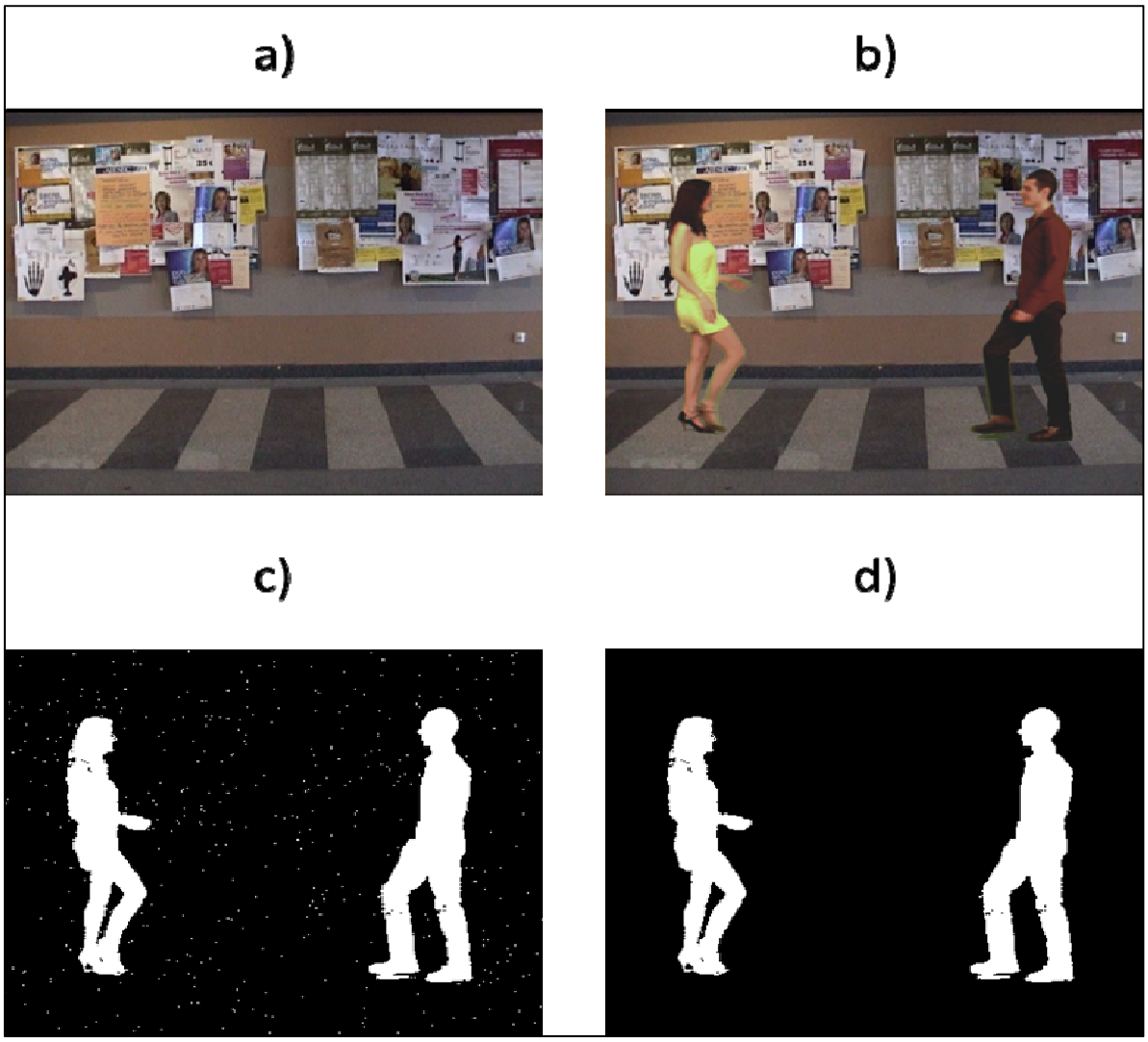


Figura 3-8: Ejemplo de extracción de blobs.

En a) se muestra el fondo del vídeo *BaileUni.avi*. En b) se puede observar el cuadro 194 del vídeo *BaileUni.av*i. En c) se muestra la segmentación del cuadro obtenida. En d) se tiene la máscara que resulta de sacar las componentes conexas de c) y filtrar por tamaño.

Se pueden observar dos blobs que se corresponden a las dos personas que forman el frente de la imagen y cómo se ha eliminado el ruido impulsivo introducido en el proceso de captación mediante filtrado por tamaño mínimo de la segmentación original.

3.7 **Modelado de frente.**

La idea es muy parecida a la del modelado de fondo. Se intenta recoger a priori toda la información del vídeo que pertenece al frente (objetos en movimiento). Para realizar el modelado del frente se suele utilizar información a nivel de blobs o de regiones como hace

J.Gallego en [25]. Los sistemas que utilizan modelado de frente para la segmentación suelen apoyarlo con esquemas clásicos de modelado de fondo con el objetivo de reducir las posibles áreas del cuadro con las que comparar el modelo de frente.

El objetivo principal de estas técnicas de modelado es obtener una imagen que recoja toda la información disponible de los objetos de frente para comparar contra ella la imagen de entrada del vídeo y poder detectar el frente de la imagen de entrada. En este punto hay que tener en cuenta que el frente puede variar de posición y de apariencia, variabilidad a considerar a la hora de realizar el modelo de frente.

La variabilidad asociada a los objetos de frente puede ser debida entre otras causas: al movimiento aparente de cada objeto por la escena, al posible alejamiento o acercamiento del objeto a la cámara o a la posible rotación del objeto de frente.

Teniendo en cuenta todas estas limitaciones, si la imagen de entrada se corresponde con el modelo esperado se actualizarán los parámetros del modelo de frente y se utilizará el modelo actualizado para el siguiente cuadro.

Enumerando las técnicas existentes que hacen uso del modelado de frente para mejorar los resultados finales de segmentación, hemos de considerar que muchos de estos algoritmos realizan un seguimiento de blobs entre cuadros consecutivos para tener una correspondencia entre los distintos cuadros [25], [29] y [31].

Se puede utilizar diversas características para realizar el seguimiento como puede ser el color o el tamaño de blob que se utiliza en el filtro de Kalman [42], el movimiento, bien el movimiento de blobs entre cuadros, o estimar el modelo de frente heredando conocimiento de observaciones anteriores [43].

Si somos capaces de estimarlo correctamente, el modelo de frente ofrece grandes beneficios entre los que cabe destacar la no pérdida de objetos de frente debidos al camuflaje de estos objetos con el modelo de fondo.

4 Procesamiento interno en cada capa.

4.1 Parámetros internos a la capa.

Existen tres tipos de parámetros que definen cada modo modelado en cada capa. El primer tipo son los parámetros que mantienen la apariencia del modo. Un segundo tipo son los parámetros que almacenan la evolución de las distancias resultantes de comparar nuevas muestras con el modo. Y finalmente existe un tercer tipo que determina la robustez del modelo del modo.

- La apariencia del modo se describe mediante un vector de media $\vec{m}_t$ valores (R, G, B) y mediante una matriz de covarianza Σ_t (varianza y correlación entre canales). Las ecuaciones de inicialización y actualización de estos parámetros se describen en detalle en el Anexo A.
- La evolución de las distancias se almacena mediante una *fdp* Gaussiana descrita por su media μ_t y desviación típica σ_t .
- Se tiene un parámetro denominado confianza *C* que mide la robustez del modelo que se describe con mayor detalle en la sección 5.2.1 y se actualiza e inicializa mediante las ecuaciones descritas en el Anexo A.

Las siguientes secciones se centrarán en la descripción del proceso de discriminación interno a cada capa.

4.2 Cálculo de la probabilidad de pertenencia a una capa: $p(z_t \mid \vec{x}_t)$

En este capítulo se describe el método de discriminación que se ha implementado en este proyecto fin de carrera para asociar cada una de las muestras de entrada $\vec{x}_t$ a alguna de las *k* capas (z_t) inicializadas en el modelo de fondo BG_t. Método que resultará en la definición de la probabilidad $p(z_t \mid \vec{x}_t)$ definida en la ecuación *eq 3.10*.

En el transcurso de este proyecto final de carrera se ha observado que esta discriminación es tan importante o más que el modelo de fondo utilizado. Para tener una medida objetiva del parecido entre las muestras de entrada y el modelo, se ha utilizado la distancia de Mahalanobis que devuelve la distancia de los tres canales de color de las imágenes y del modelo en una única dimensión. En cada píxel de cada capa se modela la evolución de la distancia de Mahalanobis resultante de comparar la muestra y el modelo para poder realizar una umbralización automática basada en los valores observados.

Idealmente valores bajos de la distancia indicarían alta similitud al modelo. Lamentablemente, existen píxeles que evolucionan en alguno de sus modos de manera ruidosa, por tanto, el uso de umbrales globales para todo el esquema de discriminación podría no ser válido para todas las muestras del modelo.

En base a esta idea, mediante el modelado de la distancia, permitimos esquemas flexibles de umbralización que se adaptan a las características observadas en el píxel. Por tanto se eliminan los umbrales globales, consiguiendo umbrales adaptados a cada pixel y a cada vídeo y no siendo necesario cambiar los umbrales para cada vídeo que se desee segmentar. Esta es una de las ventajas del algoritmo de segmentación que se presenta en este proyecto fin de carrera. El algoritmo consigue unos buenos resultados de segmentación (presentados en el capítulo 7) sin que sea necesario que el usuario inserte a priori, y tras observación empírica, el umbral óptimo de cada vídeo que se desee segmentar.

4.3 **Distancia de Mahalanobis.**

En el sistema *Bayesiano Básico* se ha considerado que los tres canales de color son independientes y por tanto se ha utilizado una matriz de covarianza diagonal para realizar el modelado. Para definir el color de cada muestra, se ha utilizado el esquema cromático *RGB*. La base teórica de esta distancia se ha contado en 3.5.5.

Para observar la mejora de añadir la correlación entre variables, primero se va a utilizar una matriz diagonal que no tiene en cuenta esta correlación (*Bayesiano Básico*) y

después se va a utilizar la matriz de covarianza completa que si las tiene en cuenta (definida en la sección de mejoras 6.2).

4.3.1 **Modelado de la distancia.**

Cuando se tiene un modelo con más de una dimensión, como puede ser el caso de un modelo de fondo con tres canales de color, se necesita modelar las diferencias por medio de más de una Gaussiana (una por cada canal de color y posterior mezcla de expertos 3.5.5). La alternativa por la que se ha optado en este proyecto ha sido trasladar las diferencias en los tres canales a una sola dimensión por medio de la distancia de Mahalanobis y modelar por medio de una *fdp* esta distancia (D). En particular, se modela por medio de una distribución Gaussiana de media μ_t y desviación típica σ_t el valor de esta distancia.

Para cada píxel del modelo de fondo BG_t, se mantiene un vector de color medio $\vec{m}_t$ y una matriz de covarianza que almacena las correlaciones entre los canales de color Σ_t. Esto significa que la media tiene tres valores, uno por cada canal de color (*RGB*) y la matriz de covarianza 9 posiciones. A lo largo de todo este proyecto se utilizará la distancia de Mahalanobis para calcular en un instante t la distancia (D_t) entre los valores almacenados en la capa ($\vec{m}_t$, Σ_t) y los valores del píxel de entrada $\vec{x}_t$.

4.3.2 **Adaptación de parámetros.**

Para adaptar los parámetros del modelado y no utilizar un umbral fijo para todos los píxeles de la imagen, se ha utilizado el método de la media móvil (ver sección 3.5.2) para ir adaptándose al sistema.

$$\mu_{t+1} = \alpha D_t + (1-\alpha)\mu_t$$
$$\sigma_{t+1} = \alpha\left((D_t - \mu_t)/K\right) + (1-\alpha)\sigma_t \qquad \textit{eq 4.1}$$

Queremos resaltar que, en la línea de esta propuesta, en este trabajo fin de carrera se ha intentado implementar un algoritmo con el menor número de umbrales y que los que finalmente se han utilizado son internos del sistema, es decir, que no requieren configuración del experto. Esto ha generado numerosos problemas ya que algunos vídeos

pueden requerir umbrales más altos que otros para conseguir los mejores resultados de segmentación. En el sistema se ha intentado que los umbrales se ajusten automáticamente a los datos de entrada, por ello se han realizado todas las pruebas del capítulo 7 con umbrales internos del sistema e **iguales** para todos los vídeos.

Los valores de actualización del parámetro α dependerán de la capa de trabajo, es decir, serán dependientes de la medida de *C* de la capa modelada (sección 7.2).

4.4 **Umbralización automática de la distancia.**

La detección de valores del píxel modelados en la capa se realiza calculando la probabilidad de que la distancia a la capa del píxel de entrada (D_t) pertenezca a la Gaussiana que modela la distancia al modelo (D) obtenida en cuadros anteriores.

La *fdp* Gaussiana que modela esta distancia presenta valores residuales a K veces el valor de la desviación de la función. Por este motivo, si la diferencia entre la media de la distancia modelada μ_t y la distancia D_t difiere en más de K veces el valor de la desviación, se considera que el valor del píxel no ha sido previamente modelado, y se determina que el píxel tiene una apariencia distinta al modo modelado (no pertenece a la capa). Mientras que, si sucede lo contrario, será considerado una nueva muestra de la apariencia modelada y se utilizará para la actualización de la capa.

$$\begin{aligned} |D_t - \mu_t| \geq K\sigma_t &\Rightarrow no\ \ pertenece\ \ a\ \ la\ \ capa \\ |D_t - \mu_t| < K\sigma_t &\Rightarrow \quad\ \ pertenece\ \ a\ \ la\ \ capa \end{aligned} \qquad \textit{eq 4.2}$$

A continuación se muestra una figura que representa gráficamente el modelado de la distancia descrito en este capítulo:

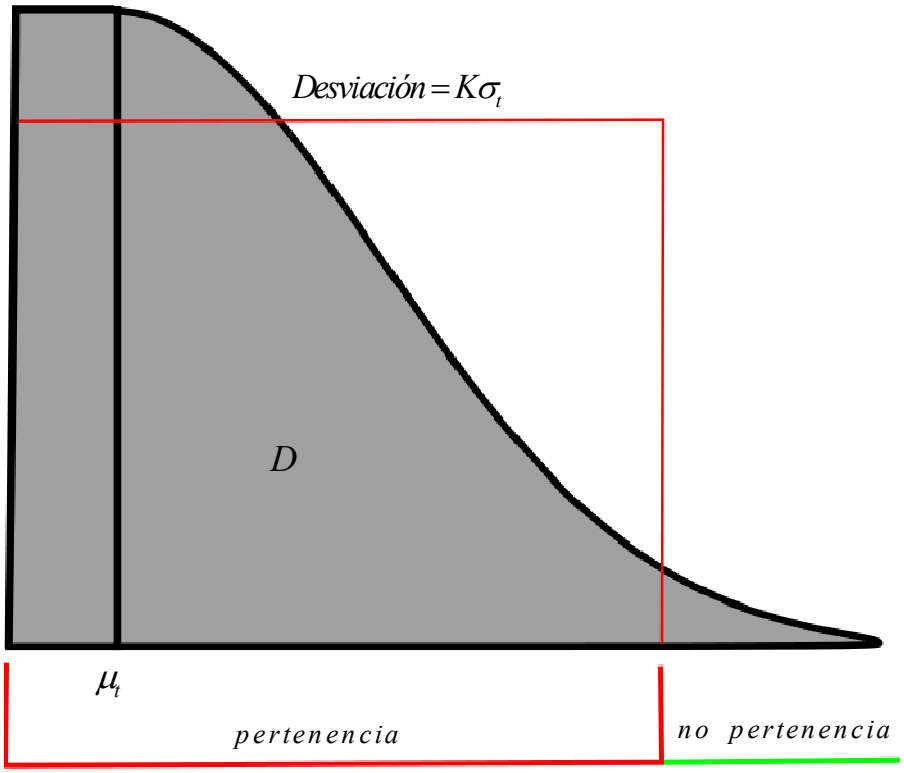


Figura 4-1: Modelado de la distancia.

5 Diseño básico. Modelado Bayesiano del Fondo.

5.1 Cálculo de la probabilidad de pertenencia al modelo de fondo: $p(BG_t \mid z_t, \vec{x}_t)$

En este capítulo se ha intentado definir la probabilidad de que dada una muestra de entrada $\vec{x}_t$, ésta pertenezca al modelo de fondo *BG* en el instante de tiempo *t*: $p(BG_t \mid z_t, \vec{x}_t)$. Para ello se ha implementado una aproximación que modela el fondo mediante un esquema bayesiano multicapa, que denominaremos *Bayesiano Básico*. El trabajo está basado en el descrito en [1].

5.2 Arquitectura del sistema.

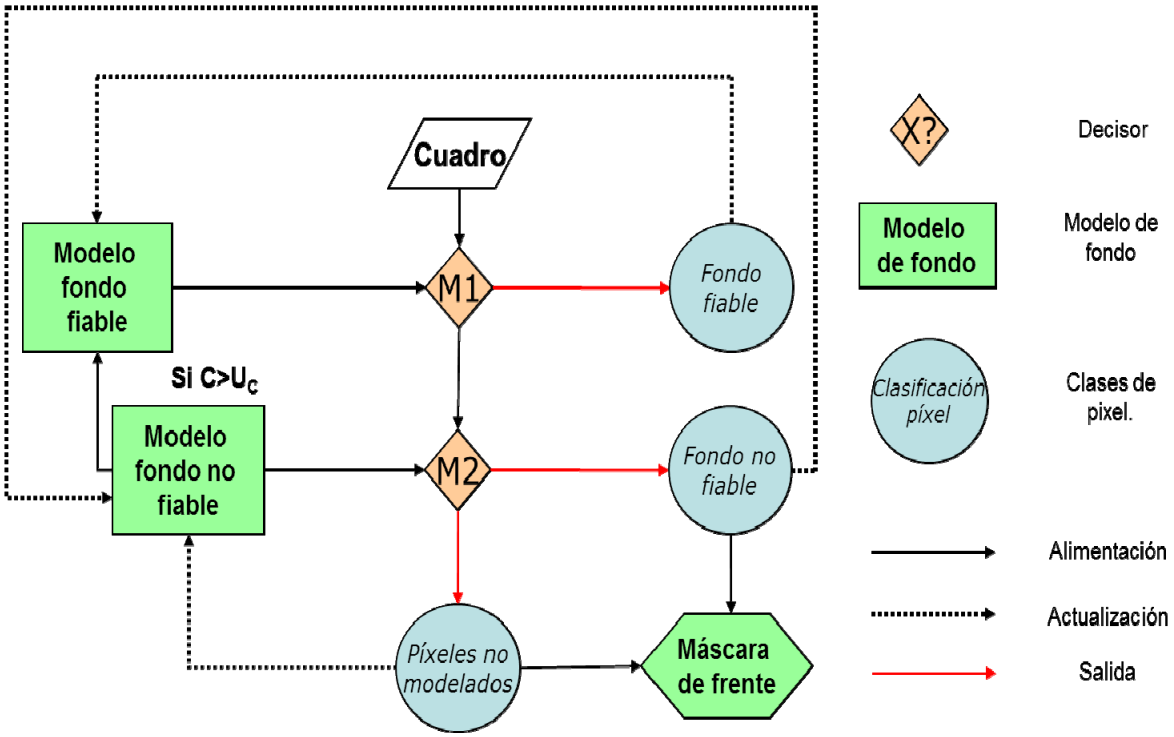


Figura 5-1: Esquema del sistema *Bayesiano Básico*.

5.2.1 Descripción de la arquitectura del sistema.

Cuadro.

Es la entrada al sistema. Es el conjunto de los píxeles de cada instante temporal del vídeo. Se compara con el modelo de fondo para determinar si corresponden a apariencias del fondo (*PoBG*) o del frente (*PoFG*).

Confianza (*C*).

La confianza es un parámetro interno a la capa que se modifica y evalúa en comparativa con las otras capas, es decir entre capas. Puede entenderse como una medida que indica la fiabilidad del modelo de capa. Su valor depende de varios factores, pero en general su valor aumenta mediante relación de proporcionalidad directa con la similitud de las muestras de entrada al modelo. Se explica con mayor detalle en 5.3.1.

Clases de pixel.

En este sistema se definen tres clases de pixel: *pixel de fondo fiable*, *pixel de fondo no fiable* y *pixel no modelado*.

- *Pixel de fondo fiable*: se clasifican en esta clase todos los píxeles que pertenecen al modelo de fondo fiable. Todos estos píxeles se marcan como fondo en la máscara final de la segmentación. (*Máscara=0*)

- *Pixel de fondo no fiable:* estos píxeles pertenecen a una capa del modelo de fondo no fiable. Se marcan como frente en la máscara final de segmentación. (*Máscara=1*). Dentro de esta clase se pueden encontrar tanto píxeles que realmente pertenecen al fondo pero que todavía no han ganado la confianza necesaria (están siendo modelados), como píxeles de frente. En este sistema los nuevos modos necesitan un tiempo de inicialización, periodo durante el cual la máscara final puede contener errores de clasificación.

- *Pixel no modelado*: son píxeles que no pertenecen a ninguna capa del modelo de fondo. Se marcan como frente en la máscara final. (*Máscara=1*). Se crea una nueva capa en el modelo de fondo no fiable con el valor del pixel. Dependiendo de las veces que se repitan valores cercanos a esta muestra durante el vídeo, esta capa pasará al modelo fiable (con confianza alta) o quedará en el modelo de fondo no fiable (con confianza baja).

Modelo de fondo.

Está compuesto de *k* capas disponibles para modelar el fondo, es decir, el sistema es capaz de almacenar *k* apariencias o modos del píxel distintos, *k* es un parámetro configurable del sistema.

Dentro del conjunto de capas de fondo se hace una distinción entre fondo fiable, bien modelado (con confianzas altas) y fondo no fiable (con confianza más baja) que se está modelando. Se toma el umbral de confianza, U_C para discriminar entre fondo fiable y fondo no fiable. Los modelos de fondo en cada capa, guardan, como se ha mencionado en la sección 4.3.1, los valores medios del píxel $\vec{m}_t$, la matriz de covarianza Σ_t , así como los valores estadísticos también descritos en esa sección (media μ_t y desviación típica σ_t) que modelan la distancia asociada a cada píxel del modelo para cada uno de los modos, (uno por capa) que han aparecido hasta el momento en el vídeo.

$$Si \longrightarrow C > U_C \longrightarrow Fondo\ fiable$$
$$Si \longrightarrow C \leq U_C \longrightarrow Fondo\ no\ fiable \qquad \textit{eq 5.1}$$

Módulo M1 y M2.

Son dos módulos de funcionamiento idéntico pero con distinta alimentación que determinan si los píxeles de la imagen de entrada pertenecen al modelo de fondo.

La diferencia fundamental entre los dos módulos es que el primero utiliza para la comparación interna a la capa las capas de fondo que tienen confianzas altas y el segundo las capas que tienen confianzas bajas (aquellas que están en proceso de modelado).

En estos módulos se efectúan las operaciones de medida de similitud del píxel de entrada con el modelo de cada capa (capítulo 4). Para aquellos píxeles que se determinen como nuevas muestras de la apariencia modelada, se determina si las capas donde se modelan pertenecen al *fondo fiable* (de apariencias *modeladas*) o pertenecen a un objeto en movimiento o un fondo en proceso de modelado (*fondo no fiable*). Aquellos píxeles que no se determinen como nuevas muestras de las capas existentes (es decir que no pertenezcan a ninguna capa), pasan a clasificarse como *pixel no modelado*.

Máscara de frente.

Es una imagen binaria que muestra el resultado final del sistema. En ella se separan los píxeles pertenecientes al fondo (*PoBG*) (*máscara=0*) y los píxeles pertenecientes al frente (*PoFG*) (*máscara=1*). Los píxeles pertenecientes al frente recogen los objetos en movimiento (que suelen ser los objetos de interés en la mayoría de las aplicaciones en las que se realiza segmentación de video).

5.3 **Modelado de fondo.**

El sistema introduce en el modelo de fondo todos los modos que adopta el píxel a lo largo del vídeo. Cada uno de los distintos modos se modela en una capa independiente. Es decir cuando se procesa una imagen, cada píxel se somete a un proceso de pertenencia comparándolo con las capas modeladas:

- Se ordenan previamente las capas en función de su confianza. La probabilidad de que las nuevas muestras pertenezcan a alguna de las capas inicializadas se estima mediante la medida de confianza.

- El proceso de pertenencia (capítulo 4) se realiza ordenadamente desde la capa con mayor confianza hasta la de menor confianza, es decir se comienza a buscar por la capa que modela el modo más repetido (de mayor probabilidad de aparición).

- En función de la pertenencia de la muestra a la capa, se adaptan los parámetros para ajustar la probabilidad de que nuevas muestras pertenezcan a dicha capa (su confianza). Adicionalmente se actualizan los parámetros internos a la capa (capítulo 4 y Anexo A).

- Si la nueva muestra no pertenece a ninguna de las capas modeladas, se inicializa una nueva capa con el valor actual del píxel y una probabilidad de que se vuelva a repetir baja. Si todas las capas están inicializadas, se elimina la de menor probabilidad de aparición (menor confianza) y se sustituye por la nueva capa.

Todas las ecuaciones que describen los procesos de adaptación e inicialización de cada capa del modelo y su desarrollo teórico están basadas en el artículo [1] de Fatih Porikli y se desarrollan en detalle en el Anexo A.

5.3.1 **Actualización de la confianza**

Siguiendo una aproximación intuitiva, la medida de confianza aumenta cuando se van encontrando más muestras que pertenecen al modo y éste se va modelando con ellas.

- **Incremento de la confianza**: cuando se encuentra pertenencia de la nueva muestra a la capa, se adaptan los parámetros aumentando la probabilidad de que nuevas muestras pertenezcan a esta capa.

- **Decrece su confianza**: todas las capas de confianza superior a la de pertenencia, donde no se ha encontrado el parecido necesario con la muestra se adaptan para reducir la probabilidad de que las nuevas muestras pertenezcan a estas capas.

Confianzas altas indican que la probabilidad de que las muestras pertenezcan a la capa es alta. Cuando una nueva capa se inicializa, la medida de confianza es baja ya que por un lado el modelo no está bien estimado porque solo se tiene una muestra y por otro lado la probabilidad de repetición es pequeña porque hasta el momento solo ha aparecido una muestra con apariencia ese modo.

Por otro lado, está muy relacionada con la varianza de la distribución del modelo ya que la medida de confianza es alta para valores pequeños de la varianza y es baja para valores altos (relación de proporcionalidad inversa).

- Cuando la varianza es pequeña, las muestras están muy cercas de la media y por lo tanto el modelo es fiable.

- Cuando la varianza del modelo es alta, las muestras que pertenecen al modelo están muy dispersas respecto a la media, y por tanto, la fiabilidad del modelo es baja.

Las ecuaciones de inicialización y actualización de la confianza a partir de las que se interpretan estas ideas pueden consultarse en el Anexo A.

5.4 La confianza como parámetro de discriminación frente-fondo.

El sistema *Bayesiano Básico* soporta la multimodalidad del fondo ya que modela los distintos modos que aparecen en el vídeo en distintas capas independientes. Cada muestra solo puede pertenecer a una única capa. Cada capa tiene asociada una confianza C que indica la fiabilidad del píxel como apariencia del fondo. Umbralizando (U_C) la confianza se determina cuáles de las capas modelan el fondo y cuáles no.

Todos los píxeles se utilizan para actualizar el modelo de fondo, de tal manera que las capas se van modelando y la confianza aumenta. Si el píxel pertenece a una capa que represente el *fondo fiable* (confianza alta) se marca como fondo en la máscara final. Los que pertenecen a un fondo sin confianza suficiente (*fondo no fiable*) se siguen marcando como frente en la máscara final, a la espera de que el modo gane la confianza suficiente y sea clasificado como fiable. Idealmente las muestras que pertenecen al frente se modelan en capas con confianzas bajas (*fondo no fiable*) y de esta manera no se pierden píxeles de frente.

Así pues, es el uso de esta confianza lo que estima la probabilidad de que tras comprobar la relación de pertenencia entre un píxel y el modo de una capa se determine si ese modo está representando el fondo o el frente. Es decir, la confianza es el parámetro que determina $p\left(BG_t \mid z_t, \vec{x}_t\right)$.

5.5 Conclusiones y limitaciones del sistema.

El sistema desarrollado aporta grandes ventajas a la hora de segmentar vídeos que contengan fondos multimodales ya que el modelo de fondo puede soportar varias apariencias distintas para cada pixel (tantas como capas).

Además cada modo se modela de manera independiente en diferentes capas de tal manera que lo que ocurre en una capa no afecta al resto. Cada vez que una nueva muestra entra al sistema se busca la pertenencia a cada una de las capas de manera independiente comenzando siempre por la que tiene mayor probabilidad de aparición. Cuando se detecta

que una muestra pertenece o no a una capa, se actualiza esta capa sin que esto afecte en nada al resto de capas.

Una limitación del sistema es que necesita un tiempo de inicialización (generalmente alto) para clasificar como *fondo fiable* cada uno de los modos, de tal manera que puede darse el caso de que un fondo multimodal cuyas apariencias se repitan poco durante el vídeo no gane nunca la confianza suficiente para ser *fondo fiable* y no se clasifique nunca como tal.

Además, y como se ha indicado anteriormente se supone que los modos donde se encuentran los píxeles de frente no van a ganar la confianza suficiente y van a ser clasificados como *fondo no fiable*. Por lo tanto se supone que estas muestras nunca se clasificarán como *fondo fiable*, situación que como hemos indicado depende del comportamiento del frente, de su apariencia y de su tamaño relativo a la imagen.

Esto repercutirá en la hipotética introducción de píxeles de frente en las capas que modelan el *fondo fiable*. Esto que supondría un problema en aquellos vídeos donde el frente bien por permanecer estático o por repetirse una misma apariencia (frentes homogéneos o persistentes) puede generar situaciones en las que los modos que los representan ganen confianza suficiente como clasificarse como píxeles de *fondo fiable* (marcándose como *máscara=0*).

Ambas limitaciones se intentarán solventar mediante una nueva clasificación del pixel, desarrollada en el siguiente capítulo (capítulo 6), donde se utilizarán nuevas características para la separación entre píxeles de frente y píxeles de fondo.

El objetivo es aislar los píxeles de frente que antes se introducían en el modelo de fondo y que ahora servirán para crear un modelo de frente que a su vez mejorará el proceso de discriminación frente fondo (como puede derivarse de la *eq 3.8*).

6 **Descripción de las mejoras.**

6.1 **Presentación de las mejoras introducidas.**

El sistema implementado parte del algoritmo *Básico Bayesiano* descrito en el capítulo 5. Las mejoras se organizan en tres grandes puntos:

- Mejorar la detección de frente introduciendo una matriz de covarianza completa que tenga en cuenta las correlaciones entre canales. Con esta mejora se busca una mejor discriminación en el modelo interno a la capa o intra-capa.

- Realizar una clasificación de pixel que permita mejorar el modelado de fondo, evitando que se contamine con frente. Si se evita que entre frente en el modelo de fondo, se puede eliminar el umbral de confianza necesario para tomar los distintos modos del pixel como fiables, es decir, binarizar $p\left(BG_t \mid z_t, \vec{x}_t\right)$. Los objetivos de esta mejora son:

    - Eliminar el tiempo de inicialización necesario para cada nueva apariencia que se introduce en el modelo de fondo

    - Conseguir que no se pierdan los objetos de frente que se repiten de manera continuada o permanecen estáticos durante un tiempo. Es decir aislar los píxeles de frente

- Realizar un modelado del frente (aislado en la segunda mejora) que permita recuperar píxeles de frente que se han clasificado de manera errónea por comparación con el modelo de fondo. Este error suele producirse por el camuflaje descrito en 3.3.

6.2 **Introducción de la matriz de covarianza completa.**

Para utilizar todas las propiedades de la distancia de Mahalanobis (ver sección 4.3), y tener en cuenta las correlaciones entre canales, se ha introducido en el algoritmo básico una matriz de covarianza completa.

En el sistema se ha considerado las correlaciones entre los tres canales de color de un pixel para calcular la matriz de covarianza. En concreto (y a lo largo de todo el proyecto)

hemos utilizado el espacio de color *RGB*. La matriz de covarianza utilizada tiene la siguiente forma:

$$\sum = \begin{pmatrix} \sigma_R\sigma_R & \sigma_R\sigma_G & \sigma_R\sigma_B \\ \sigma_G\sigma_R & \sigma_G\sigma_G & \sigma_G\sigma_B \\ \sigma_B\sigma_R & \sigma_B\sigma_G & \sigma_B\sigma_B \end{pmatrix} \qquad \textit{eq 6.1}$$

Donde σ_Z indica la desviación típica de cada canal.

Al introducir la matriz de covarianza completa se busca introducir una nueva variable (o característica), la distribución temporal, muestreada en cada pixel modelado, de los canales de la imagen. Esta distribución puede ayudar a la discriminación de frente fondo descrita en 3.5.5.

De este modo, mientras se respete esta distribución, es decir, una nueva apariencia del pixel con distribución similar a la del modo modelado, la matriz de covarianza apoyará distancias cercanas a las modeladas. En caso contrario, si el píxel de entrada presenta distribuciones distintas al modo modelado, al calcular la distancia de Mahalanobis utilizando una distribución distinta a la actual, se obtendrán distancias distintas a las descritas mediante la *fdp* Gaussiana que almacena la distancia (ver sección 4.3.1).

Se ha realizado una implementación de la distancia de Mahalanobis con matriz de covarianza completa utilizando ventajas del cálculo matricial para que los sistemas implementados no fuesen demasiado costosos computacionalmente. Esta implementación se describe en el Anexo B.

6.3 **Modelado de fondo utilizando clasificación de píxel.**

El sistema básico introduce en el modelo de fondo todas las apariencias que presenta el pixel a lo largo del vídeo, incluidas las que pertenecen a objetos de frente. Para no perder en el resultado final del sistema los objetos del frente que aparecen durante el vídeo, se utiliza un umbral de confianza mínimo para discriminar entre capas de *fondo fiable* y de *fondo no fiable*. Esto requiere de un tiempo de inicialización para permitir la introducción de nuevas apariencias en el modelo de fondo (mientras aumenta la confianza hasta superar el valor de umbral).

Por el contrario, con la mejora que se propone se busca evitar la introducción de apariencias de frente en el modelo de fondo para tener un sistema capaz de obtener resultados aceptables para un mayor número de vídeos. Se ha observado que la introducción de objetos de frente en el modelo de fondo es especialmente preocupante en vídeos donde el mismo objeto de frente pasa varias veces por la escena o en vídeos donde un objeto de frente permanece estático durante un cierto tiempo (incrementando la confianza de los modos que lo modelan en el sistema *Bayesiano Básico*).

En estos casos el frente consigue una confianza más alta que el umbral exigido pasando a ser *fondo fiable*. Además con la nueva clasificación de pixel se busca reducir el tiempo de conversión de *fondo fiable* a *fondo no fiable* (disminución del tiempo de inicialización de las nuevas apariencias de fondo multimodal que aparecen en las secuencias a segmentar)

El objetivo final es conseguir un sistema robusto que presente un funcionamiento similar al sistema *Bayesiano Básico* para fondos multimodales y que además sea capaz de segmentar objetos de frente que se repitan varias veces durante la duración del vídeo o que permanezcan estáticos.

La organización de esta sección es la siguiente. Inicialmente se describen las nuevas clases de píxel 6.3.1. Después, se ha planteado la mejora en dos etapas, primero se ha realizado una clasificación del píxel utilizando sólo información de bajo nivel (pixel aislado) 6.3.2, mejora con la que no se consiguió eliminar el umbral de confianza ni la división del fondo modelado en *fondo fiable* y *fondo no fiable*. Por ello, como una 'mejora sobre la mejora' se realizará la misma clasificación incorporando además información a nivel de blob (píxeles homogéneos en movimiento) 6.3.3, donde se ha conseguido la eliminación del umbral de confianza.

Cada una de estas mejoras se va a evaluar por separado (capítulo 7) para poder medir la aportación de cada una de ellas.

6.3.1 **Clases de pixel.**

Se clasifican todos los píxeles de la imagen de entrada en cinco clases: *fondo modelado, fondo dinámico sin modelar, fondo estático sin modelar, frente* y *píxeles sin clasificar*. Todos los píxeles que no están clasificados (*píxeles sin clasificar*) se meten en una capa intermedia a la espera de ser clasificados.

- ***Fondo modelado***: a esta clase pertenecen todos los píxeles que pertenecen a alguno de los modos ya inicializados en el modelo de fondo. Inicialmente en la mejora a nivel de píxel, se dividirá en las clases propuestas anteriormente en el *Bayesiano Básico*: *fondo fiable* y *fondo no fiable*.

Posteriormente, se intentarán sustituir estas clases por un única de *fondo modelado* independientemente de la confianza que tengan, que sustituiría a las clases de *fondo fiable* y *fondo no fiable*. Esta sustitución no se conseguirá hasta la introducción de información a nivel de blob.

- ***Fondo dinámico no modelado***: a esta clase pertenecen los píxeles que pertenecen a un modo no inicializado en el modelo de fondo, pero que dispone de otros modos en la misma posición en donde la medida de confianza tiene una evolución oscilante.

Como ejemplo, podemos pensar en una pared que se ve detrás de la copa de un árbol movido por el viento. El modo que describe la pared no está inicializado pero sí que lo está el modo que describe la copa del árbol y que se está repitiendo durante la duración del vídeo.

- ***Fondo estático no modelado***: son píxeles que todavía no han sido inicializados en el modelo y que además no disponen de ningún modo que gane confianza. En estos píxeles, la medida de confianza baja bruscamente en todos los modos inicializados.

La dificultad radica en que este comportamiento también es típico de los píxeles de frente. Por lo tanto es muy difícil hacer una distinción entre los dos tipos. La diferencia es que los fondos estáticos no cambian de posición con el tiempo mientras que los píxeles de frente suelen tener un movimiento asociado. A esta clase pertenecerían zonas donde ha habido un cambio brusco de iluminación o zonas donde se ha inicializado el modelo de fondo con un objeto de frente que ocluía el verdadero fondo (*inicio en caliente*).

- ***Frente***: cuando un objeto de frente pasa por delante del fondo, las medidas de confianza de todos los modos inicializados bajan de manera brusca hasta que el objeto de frente deja de pasar por delante. En el caso de frentes estáticos el comportamiento es muy parecido al del fondo estático provocando problemas para diferenciarlos. Para conseguir hacer una diferenciación, vamos a utilizar tanto a nivel de píxel como a nivel de blob la evolución de la confianza, además a nivel de blob vamos a suponer que el frente estático *es menos estático* que los fondos estáticos. Es decir vamos a suponer que los objetos de frente siempre tienen un cierto movimiento, como por ejemplo el movimiento de las extremidades en el caso de personas.

- ***Píxeles sin clasificar*:** todos los píxeles que no se clasifican como ninguno de los tipos anteriores se incluyen en una capa intermedia a la espera de asignación de clase. En este tipo se incluyen todos los píxeles que no han cumplido las reglas para pertenecer a las clases anteriores y por lo tanto no se clasifican por no disponer de los datos necesarios. Más que como una clase en sí misma, puede entenderse como un *contenedor* que evita la pérdida de información potencialmente útil a la espera de evidencias que permitan clasificar los píxeles.

6.3.2 **Utilizando información de bajo nivel (pixel aislado).**

En esta mejora se pretende que los píxeles de *frente* no se modelen en el modelo de fondo. Para ello los píxeles no clasificados como *fondo modelado (fiable y no fiable)* se clasifican en función de una nueva característica, la evolución de la confianza en las diferentes capas que el pixel tenga inicializadas en el modelo de fondo. Dependiendo de la evolución de esta medida a corto plazo, se discrimina por un lado el *fondo dinámico no modelado* y por otro el conjunto de *frente* y *fondo estático no modelado*.

La discriminación entre *frente* y *fondo estático no modelado* es una tarea compleja y por lo tanto se crea una nueva capa para introducir estas muestras y realizar un estudio que permita separar el *fondo estático no modelado* para incluirlo en el modelo de fondo, aislando así el *frente*.

Con este objetivo, se establece un ***tiempo de margen*** que corresponde al tiempo necesario para que el modo de la nueva capa creada alcance confianza necesaria para ser

clasificado como *fondo estático no modelado*. El pixel se caracterizará como tal siempre que durante este ***tiempo de margen*** la medida de confianza siga bajando bruscamente en todas las capas inicializadas en el modelo de fondo, en caso contrario, se clasificará como *frente*.

6.3.2.1 **Arquitectura del sistema.**

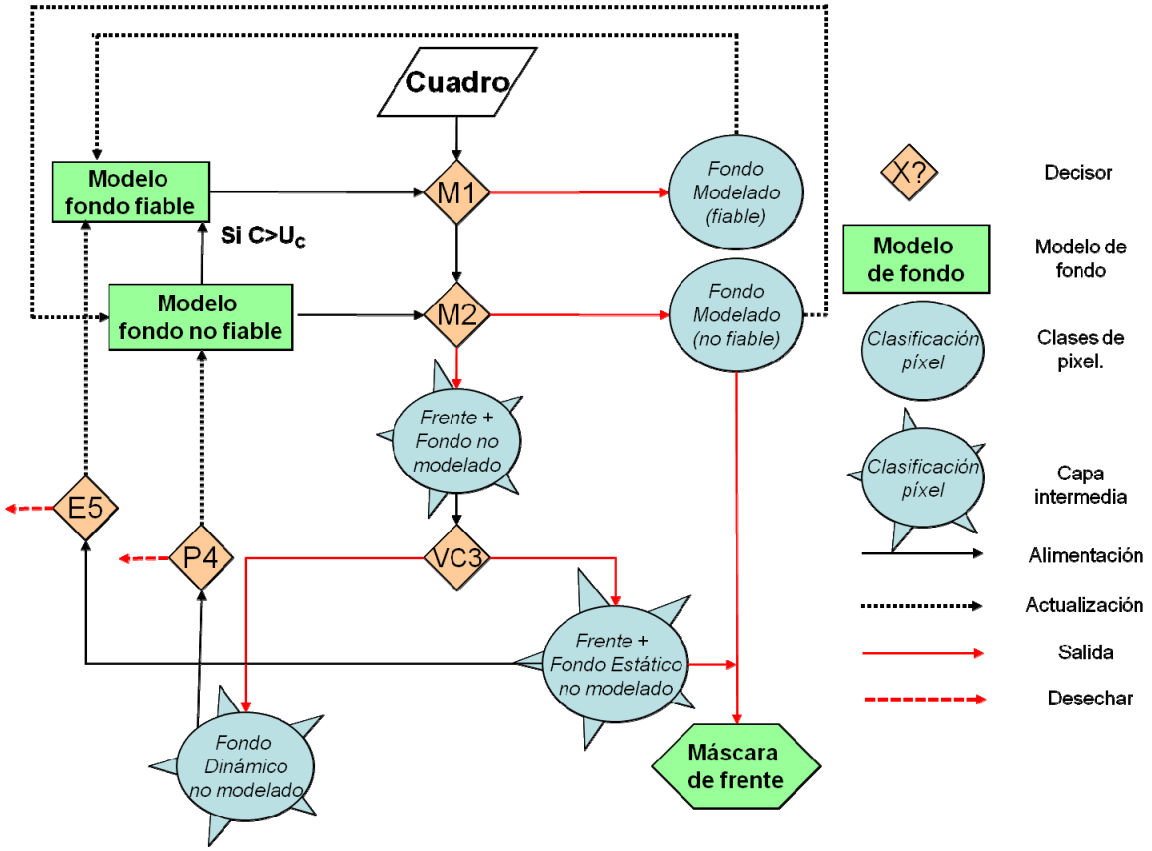


Figura 6-1: Esquema del sistema mejorado a nivel de pixel.

6.3.2.2 **Descripción de la arquitectura del sistema.**

Dentro del modelo de fondo se sigue haciendo la distinción entre *fondo fiable*, bien modelado (con confianza alta) y *fondo no fiable* (con confianza baja) que se está modelando. Solo se describen los nuevos módulos que se han introducido en el sistema. El resto ya se definieron en 5.2.1.

Módulo VC3. Modelado de la variabilidad en la confianza.

En este módulo se estudia la variabilidad de la medida de confianza para cada pixel del modelo de fondo. Se ha creado una capa intermedia para almacenar los píxeles a la espera de clasificación.

Cuando un pixel entra en la capa intermedia, se queda en ella durante los siguientes cuadros para estudiar cómo evoluciona la medida de confianza en las capas que tiene inicializadas el modelo de fondo en esta posición. Dependiendo de la evolución (ver Tabla 6-1) será clasificado en distintas clases descritas en 6.3.1.

Específicamente, en este módulo los píxeles se dividen en dos tipos dependiendo de la evolución. Si la medida de confianza baja de manera brusca en todas las capas del modelo, el pixel se clasifica como *frente* y *fondo estático no modelado*. Esta indeterminación en la clasificación es uno de los problemas derivados de esta mejora, se intentará solventar este problema en el módulo E5.

Si por el contrario la medida de confianza oscila en alguna de las capas del modelo, esto quiere decir que se están clasificando alguna de las siguientes muestras del píxel como *fondo modelado*, el pixel se clasifica como *fondo dinámico no modelado*.

En esta clase se incluyen la mayor parte de los píxeles de fondo multimodal pero lamentablemente, también se introducen algunos píxeles de frente, bien porque son partes del frente que se mueven muy rápido o porque existe camuflaje de tal manera que la confianza tiene una evolución oscilante igual que los píxeles de fondo multimodal, se intentará minimizar la intrusión de píxel en esta clase mediante el módulo P4.

A continuación se muestra una figura donde se puede observar a) la evolución de la confianza a lo largo de los primeros 210 cuadros de una secuencia de prueba para un pixel $\vec{x}_{1:210}$. b) la evolución de la medida de confianza para un pixel de *frente* entre los cuadros 80 y 85 de la secuencia de prueba y c) una evolución típica de *fondo dinámico no modelado* entre los cuadros 5 y 8.

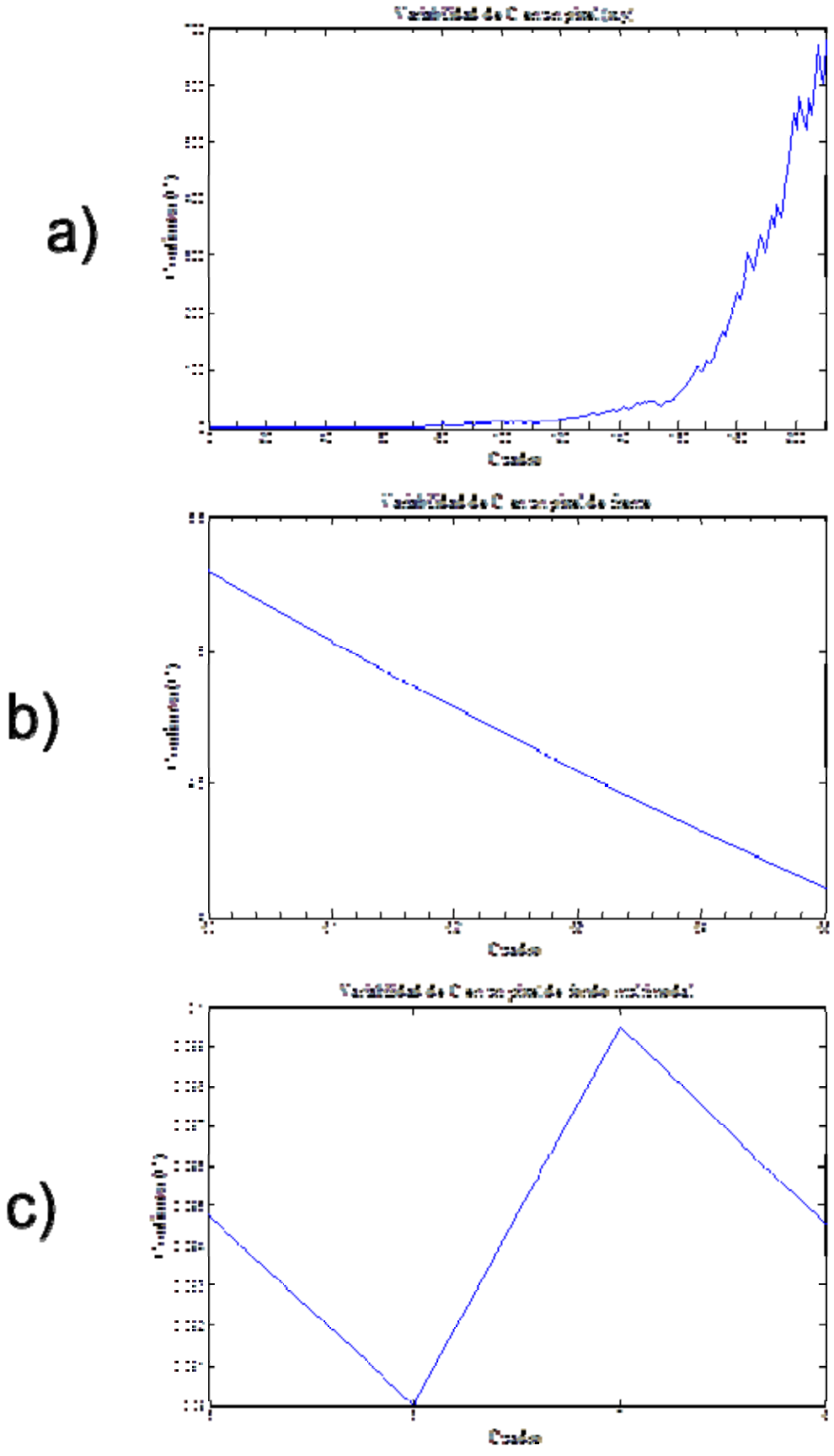


Figura 6-2: Variabilidad de la confianza.

Módulo P4. Busca repeticiones periódicas.

Este módulo se añade porque se ha comprobado que teniendo en cuenta sólo la variabilidad de la confianza se introduce algún pixel de *frente* como *fondo dinámico no modelado*. Es por lo tanto un módulo depurador más que un módulo clasificador.

En él, se buscan repeticiones periódicas ya que se supone que si realmente es fondo multimodal se va a repetir el mismo modo en alguno de los siguientes cuadros del vídeo. También se supone que el pixel de *frente* (que ha llegado hasta aquí por los problemas derivados de trabajar a nivel de pixel) se va a desplazar y no va a aparecer en los siguientes cuadros.

Siguiendo una solución empírica, el módulo estudia el pixel que queremos introducir en el modelo de fondo durante tres periodos de diez imágenes. Por tanto, para clasificar el pixel bajo estudio como *fondo dinámico no modelado* e introducirlo en el modelo de fondo necesitamos encontrarlo en alguno de los diez cuadros de cada periodo. Es decir, buscamos que el nuevo pixel a introducir en el modelo se repita '*periódicamente*' (con periodo indeterminado pero menor a 10). Situación que se correspondería con la evolución estándar de la mayoría de los fondos multimodales o dinámicos.

Módulo E5. Extractor de fondo estático

En este módulo se intenta separar el *frente* del *fondo estático no modelado* que se encuentran mezclados ya que la medida de confianza baja bruscamente en los dos casos.

Utilizando solo información a nivel de pixel esta separación es muy difícil ya que un pixel clasificado como *fondo estático no modelado* y algunos pixeles de *frente* (sobre todo si éste permanece estático o presenta apariencias repetitivas) tienen comportamientos prácticamente iguales. Para poder recuperar las zonas de *fondo estático no modelado* se han introducido todos los píxeles con bajada brusca de la confianza en una nueva capa intermedia que comparamos y actualizamos con las nuevas imágenes de entrada.

Los píxeles que no pertenecen nunca a los modelos de fondo y que tienen confianzas muy altas dentro de la capa intermedia se clasificarán como *fondo estático no modelado* y se introducirán en el modelo de fondo. Hay que destacar que empíricamente se ha observado que mediante esta estrategia se necesita un **tiempo de margen** elevado (muchas

muestras del píxel) para poder clasificarlo como *fondo estático no modelado*. Incluso así se siguen introduciendo en el modelo de fondo pixeles de *frente* que han permanecido estáticos durante un cierto tiempo.

Para mejorar esta separación y conseguir un modelo de *frente* más fiable que permita mejorar los resultados de la segmentación final, se va a continuar el proyecto utilizando información a nivel de blob. Esta información es más robusta puesto que aprovecha la correlación espacial de las muestras eliminando la inestabilidad en la clasificación del pixel. Permitiendo así la clasificación de un conjunto de píxeles en vez de la de píxeles aislados. Gracias a esta nueva información se logra una separación robusta entre *frente* y *fondo estático* que no se consigue en este módulo (E5).

6.3.2.3 **Resumen de las características analizadas.**

La siguiente Tabla resume las características analizadas para realizar la clasificación de pixel en la aproximación a bajo nivel o nivel de píxel.

Clase	**Evolución de la Confianza a corto plazo**	**Evolución de la Confianza durante el tiempo de margen**
fondo modelado	Irrelevante	Irrelevante
fondo estático no modelado	Bajada de la confianza brusca en las capas que el pixel tiene inicializadas en el modelo de fondo	Bajada de la confianza brusca en las capas que el pixel tiene inicializadas en el modelo de fondo
frente	Bajada de la confianza brusca en las capas que el pixel tiene inicializadas en el modelo de fondo	Valores alternativos de bajada y subida de confianza
fondo dinámico no modelado	Valores alternativos de bajada y subida de confianza	Valores alternativos de bajada y subida de confianza

Tabla 6-1. Clasificación de pixel en función de la variabilidad de la confianza.

6.3.3 **Utilizando información a nivel de blob.**

Se utiliza este nivel de información definido en 3.6 para mejorar la clasificación a nivel de pixel que se ha explicado en la sección anterior 6.3.2. Ahora la clasificación se realiza utilizando información de una región de píxeles homogéneos en movimiento en lugar de realizarla sobre píxeles aislados. De esta forma se ha observado que la clasificación es mucho más robusta. Además se dispone de nuevas características para la clasificación como el movimiento de los blobs, su forma…etc. En definitiva, al trabajar a nivel de blob se tiene información espacial que antes no se tenía.

Los objetivos son mejorar la clasificación de pixel descrita en la sección 6.3. Se quiere introducir en el modelo de fondo un mayor número de píxeles de fondo multimodal (*fondo dinámico no modelado*), píxeles que se perdían utilizando únicamente información de bajo nivel.

De la misma forma, también se busca minimizar el número de píxeles incorrectamente clasificados como *fondo dinámico no modelado*, cuando en realidad eran *frente*, píxeles que se introducían en el modelo de fondo, contaminándolo, en la aproximación anterior.

Además, con las nuevas características para la clasificación se intentará separar los píxeles de *frente* de los *fondos estáticos no modelados* de una manera más eficaz ya que como se ha visto anteriormente esta clasificación no era muy precisa con la información de bajo nivel. Gracias a ello, será más fácil resolver problemas clásicos de los algoritmos de segmentación como son los *inicios en caliente* (objetos de frente en las primeras imágenes del vídeo), o modelado de los cambios bruscos de iluminación. También se utiliza esta información para realizar un modelado de los píxeles clasificados como *frente*, modelo que se detalla en la sección 6.4 y servirá para clasificar correctamente píxeles de frente que se *camuflan* con el fondo.

Por último se intentará reducir o eliminar el número de muestras necesario para la inicialización y aprendizaje de las nuevas apariencias que entran en el modelo de fondo. Se va a intentar que todo lo que entre en el modelo de fondo se clasifique como *fondo modelado*, eliminando el *fondo no fiable* del algoritmo original y por tanto el umbral

exigido de confianza. Con esta mejora se obtiene una máscara de segmentación desde la segunda imagen del vídeo procesado, sin tener que esperar el tiempo de inicialización requerido tanto en el sistema *Bayesiano Básico* como en la aproximación de bajo nivel descrita en la sección anterior.

Obviamente, para el correcto funcionamiento del sistema, y debido a la eliminación del umbral de confianza, se priorizará la no introducción de muestras de *frente* en el modelo de fondo.

6.3.3.1 **Arquitectura del sistema.**

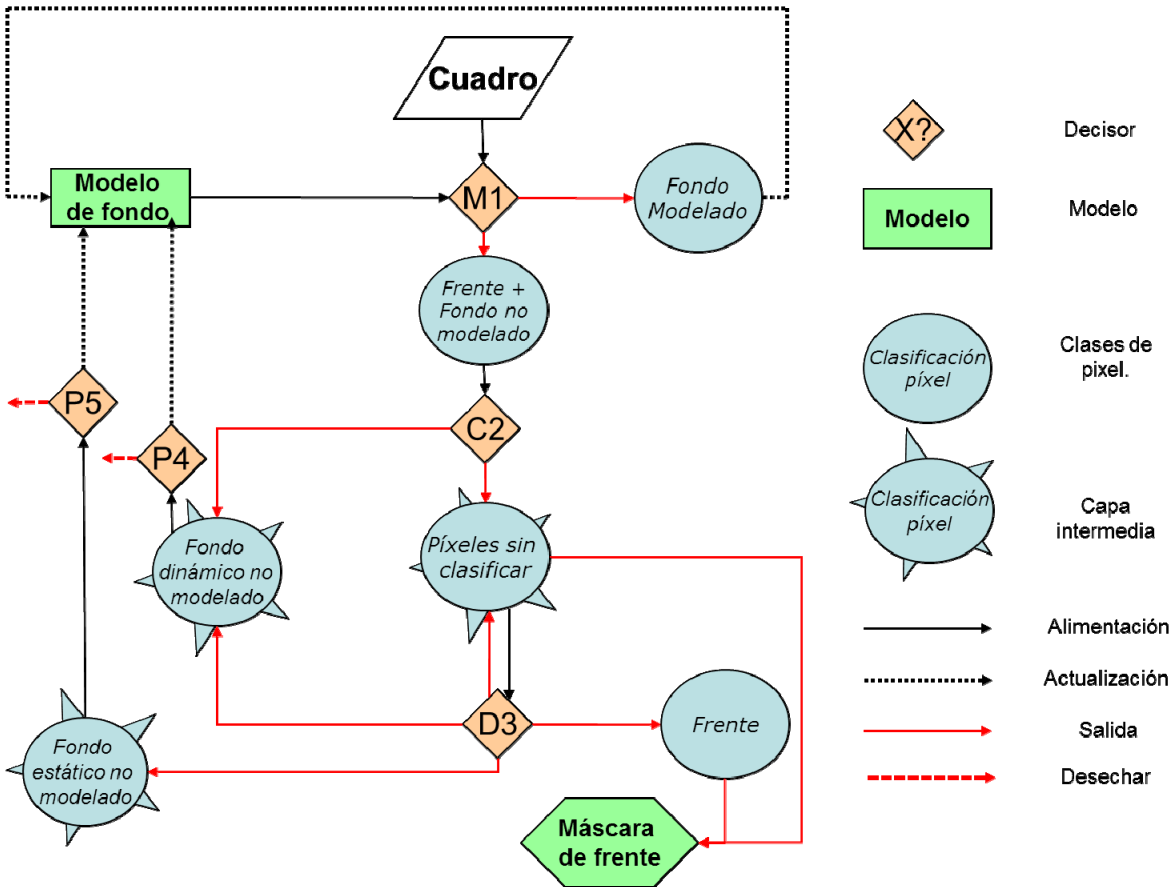


Figura 6-3: Esquema del sistema mejorado a nivel de blob.

6.3.3.2 **Descripción de la arquitectura del sistema.**

Con la nueva clasificación de pixel que se ha realizado utilizando información a nivel de blob, se cometen menos errores en la clasificación final. De esta manera se tiene un mejor modelo de fondo y se puede aislar el *frente* que modelándolo posteriormente

permitiría mejorar los resultados finales de la segmentación. Adicionalmente se elimina el umbral de confianza utilizado en los sistemas anteriores reduciendo con ello el tiempo de inicialización de nuevas apariencias de fondo del algoritmo original y tomando todo el modelo de fondo como *fondo modelado*. En este sistema se fusionan las clases *fondo fiable* y *fondo no fiable* en esta nueva clase denominada *fondo modelado*. Solo se describen los nuevos módulos que se han introducido en el sistema. El resto ya se definieron en 6.3.2.2.

Módulo M1. Comparación con el modelo de fondo.

En este módulo se calcula la distancia de los píxeles de entrada del cuadro con los que se tienen almacenados en el modelo de fondo. Si la apariencia se ajusta a alguno de los modos almacenados en el modelo de fondo (ver capítulo 4), el pixel se clasifica como *fondo modelado* independientemente de la confianza del modo y se marca como fondo en la máscara final de la segmentación. El resto de píxeles que no se ajustan a ninguna de los modos modelados se clasifican como *frente* o *fondo no modelado* y se introducen al módulo C2 para seguir con su tratamiento. Es una evolución del módulo M1 definido en 5.2.1.

Módulo C2. Extrae las componentes conexas para formar blobs.

Extracción de blobs.

Todos los píxeles que no pertenecen a *fondo modelado* se pasan a este módulo que forma blobs con las componentes conexas y elimina los píxeles aislados de la máscara de frente (filtrado por tamaño).

Para extraer los blobs con los que posteriormente se realiza la clasificación de pixel se ha modificado un algoritmo disponible en el grupo de investigación *VPULab* que busca componentes conexas en una imagen y las etiqueta con un número característico (*BlobId*) si cumple un requisito configurable de tamaño mínimo. Si la componente es menor que el tamaño definido se clasifica como *fondo dinámico no modelado* y se elimina de la máscara de blobs.

Cada uno de los blobs extraídos queda caracterizado en el instante *t* por: su *BlobId*, su máscara ($Máscara_t$), la altura y anchura del rectángulo que lo contiene (H_t, W_t), y las

coordenadas de su punto central (x_t^{PC}, y_t^{PC}). Cada uno de éstos parámetros se indican en la Figura 6-4 para el mismo blob (misma *BlobId*) en dos instantes: *t* y *t-n*.

Figura 6-4: Caracterización del blob.

Mediante el proceso de filtrado por tamaño descrito es posible eliminar falsos positivos resultantes de pequeñas variaciones del píxel no modeladas. Estas variaciones pueden ser producidas por el ruido impulsivo captado en el proceso de grabación, el propio proceso de grabación (p.e. *jitter* de la cámara) o por píxeles aislados de fondo multimodal.

Todos estos píxeles que no pertenecen al modelo de fondo y que además no superan el filtrado por tamaño, se clasifican como *fondo dinámico no modelado* y se introducen en P4 para su tratamiento. Estos píxeles también se marcan como fondo en el resultado final de la segmentación.

El resto de píxeles los que han superado el filtrado por tamaño, se incluyen en una capa intermedia a la espera de ser clasificados. Todos estos píxeles se marcan como frente en la máscara final.

Módulo D3. Módulo decisor. Clasifica los blobs.

En la capa intermedia que contiene los *píxeles sin clasificar*, los blobs permanecen durante varios cuadros hasta que se tienen los datos suficientes para poder clasificarlos. Para cada cuadro de entrada, se pasa a la capa intermedia una imagen de blobs a los que se busca correspondencia mediante un algoritmo de seguimiento de blob.

Seguimiento de blobs.

Para poder estudiar las características analizadas en el sistema durante una serie de cuadros, es necesario saber qué blob corresponde a uno dado en los siguientes cuadros de entrada. Para ello el sistema utiliza un filtro de *Kalman* [42] resultado de la adaptación e integración de un algoritmo existente en el grupo de investigación *VPULab*. Dicho algoritmo toma como entrada los blobs aislados de cada cuadro, y que da como resultado el extractor de blobs descrito anteriormente en el módulo C2, donde se han filtrado por tamaño componentes conexas pequeñas y se han etiquetado los blobs.

El filtro de *Kalman* tiene una memoria configurable de *n* imágenes de tal manera que para cada blob se tienen almacenadas las últimas *n* imágenes en la que se haya encontrado correspondencia entre el blob y los blobs de entrada. Si el sistema tiene memorizado un blob para el que no se ha encontrado correspondencia en los últimos *n* cuadros, el blob se elimina del sistema.

Caracterización de los blobs.

Para todos los blobs que hay en la capa intermedia, se calculan dos nuevas características que resultan de la nueva información resultante al trabajar a nivel de blob, son: el *movimiento* (desplazamiento del blob entre cuadros) y el *solape* entre un blob con sus blobs anteriores (misma *BlobId*) en los distintos cuadros donde se ha detectado y que se tienen en memoria.

Cálculo del solape.

Por medio del solape se tiene una idea de cómo evoluciona la medida de confianza de forma global para todos los píxeles que forman un blob.

El proceso de cálculo del solape es simple, se toman dos muestras del mismo blob (blobs en dos instantes temporales: *t* y *t*-*n*, siendo *n* un parámetro configurable, pero con la

misma *BlobId*) y se calcula la intersección entre las máscaras que los representan a ambos. Posteriormente, se calcula la unión de las dos máscaras. La medida de solape resultará de la división de ambos indicadores:

$$Solape = \frac{B_t^{BlobId} \cap B_{t-n}^{BlobId}}{B_t^{BlobId} \cup B_{t-n}^{BlobId}} \qquad eq\ 6.2$$

Puede observarse que si se fija *n=1*, estaremos estimando la evolución de la confianza que sufren los modos del modelo de fondo que modelan los píxeles que forman el blob.

Otra interpretación permitiría ver el solape, como una medida de la estabilidad de la forma del blob. Así, si la forma es exactamente igual o muy parecida entre los blobs (solape alto) podríamos clasificar todos los píxeles del blob como *fondo estático no modelado* o *frente*, mientras que si la forma ha cambiado (solape bajo) es probable que todos los píxeles del blob sean *fondo dinámico no modelado*.

Cálculo del movimiento.

El movimiento, o desplazamiento espacial del blob en el tiempo, puede ayudar a mejorar la clasificación de los píxeles que forman cada blob. Se utiliza como complemento a la medida de solape del blob. Al igual que el solape se evalúa el movimiento del punto central de un mismo blob en dos instantes temporales: *t* y *t-n*. El movimiento de un blob se estima como la distancia Euclídea entre sus puntos centrales.

$$Movimiento = \sqrt{(x_{t-n}^{PC} - x_t^{PC})^2 + (y_{t-n}^{PC} - y_t^{PC})^2} \qquad eq\ 6.3$$

Este movimiento se normalizará en función de la resolución de la imagen, pero no del tamaño del blob, normalización que se propone como trabajo futuro.

Para clasificar un blob como *frente* es necesario que presente un movimiento alto ya que se supone que un objeto de frente se tiene que mover para entrar en la escena aunque es posible que posteriormente se detenga. Por el contrario, si se detecta que un blob no ha tenido movimiento o el movimiento ha sido muy próximo a cero, es probable que el blob pueda clasificarse como *fondo estático no modelado* o como *fondo dinámico no modelado*.

Clasificación del blob.

Dependiendo del solape y el movimiento que caracterizan un blob en cada instante temporal, podemos clasificar los píxeles de dicho blob como *fondo estático no modelado*, *fondo dinámico no modelado*, *frente* o como *píxeles sin clasificar* (ver 6.3.1). En este último caso el blob permanece en la capa intermedia y se vuelve a evaluar en la siguiente iteración con la nueva información obtenida. Si por el contrario el blob se ha clasificado, se extrae de la capa intermedia eliminando toda la información que se tenía de él.

Los parámetros de clasificación de un blob son:

- Si su movimiento es alto y su solape alto, se clasifica como *frente*.
- Si su movimiento es bajo y su solape bajo, se clasifica como *fondo dinámico no modelado*
- Si su movimiento es bajo y su solape alto, se clasifica como *fondo estático no modelado*.
- En caso contrario permanece en la capa intermedia y se marca como *píxeles sin clasificar*.

La definición de 'alto' y 'bajo' depende de la naturaleza del video a analizar, se han fijado valores empíricamente, pero se han mantenido iguales para todos los videos analizados en el capítulo de resultados (capítulo 7).

Cuando todas las características indican que el blob pertenece a una misma clase, se clasifican todos los píxeles que forman el blob como muestras de dicha clase. Si, por el contrario existen contradicciones entre las distintas características utilizadas para la clasificación, se deja el blob en una capa intermedia y todos sus píxeles se marcan como *píxeles sin clasificar*, a la espera de poder mejorar la clasificación con la información contenida en los siguientes cuadros de entrada.

Módulo P5. Extractor de fondo estático.

Es una versión mejorada de módulo E5 descrito en 6.3.2.2. En este módulo se evalúa que los píxeles que han sido clasificados como *fondo estático no modelado* realmente lo son antes de introducirlos en el modelo. Para ello se supone que ningún pixel de los siguientes cuadros de entrada en estas posiciones puede pertenecer al modelo de fondo. Es decir, en estas posiciones solo han de aparecer pixeles clasificados como *frente* (que no

deberían pertenecer al modelo de fondo) o nuevas apariencias del *fondo estático no modelado* que estamos evaluando (que todavía no está inicializado en el modelo de fondo). Si esto se cumple durante los siguientes 10 cuadros, los píxeles clasificados como *fondo estático no modelado* se introducen en el modelo de fondo.

Gracias a que la nueva clasificación de pixel a nivel de blob es capaz de aislar adecuadamente los píxeles de *frente*, se puede introducir la última mejora que se ha implementado en este proyecto fin de carrera. Se trata de un modelado de frente que se añade al sistema de clasificación de pixel a nivel de blob descrito en esta sección.

6.4 **Modelado de frente.**

Ésta es la última mejora implementada. Su objetivo es recuperar a posteriori algunos de los píxeles que deberían haberse clasificado como *frente*, pero se han perdido en la máscara final de segmentación por problemas como el camuflaje (descrito en 3.3) con el *fondo modelado*.

En el esquema 6.3.3.1 se muestra como con la clasificación de pixel a nivel de blob, se consigue aislar los píxeles de *frente*. En ese esquema se aislaba el *frente* pero no se utilizaba para mejorar el resultado final del sistema, simplemente se separaban los píxeles clasificados como tal en una nueva capa.

A continuación se muestra el nuevo esquema de segmentación donde se introduce un modelo de frente para mejorar los resultados:

6.4.1 **Arquitectura del sistema.**

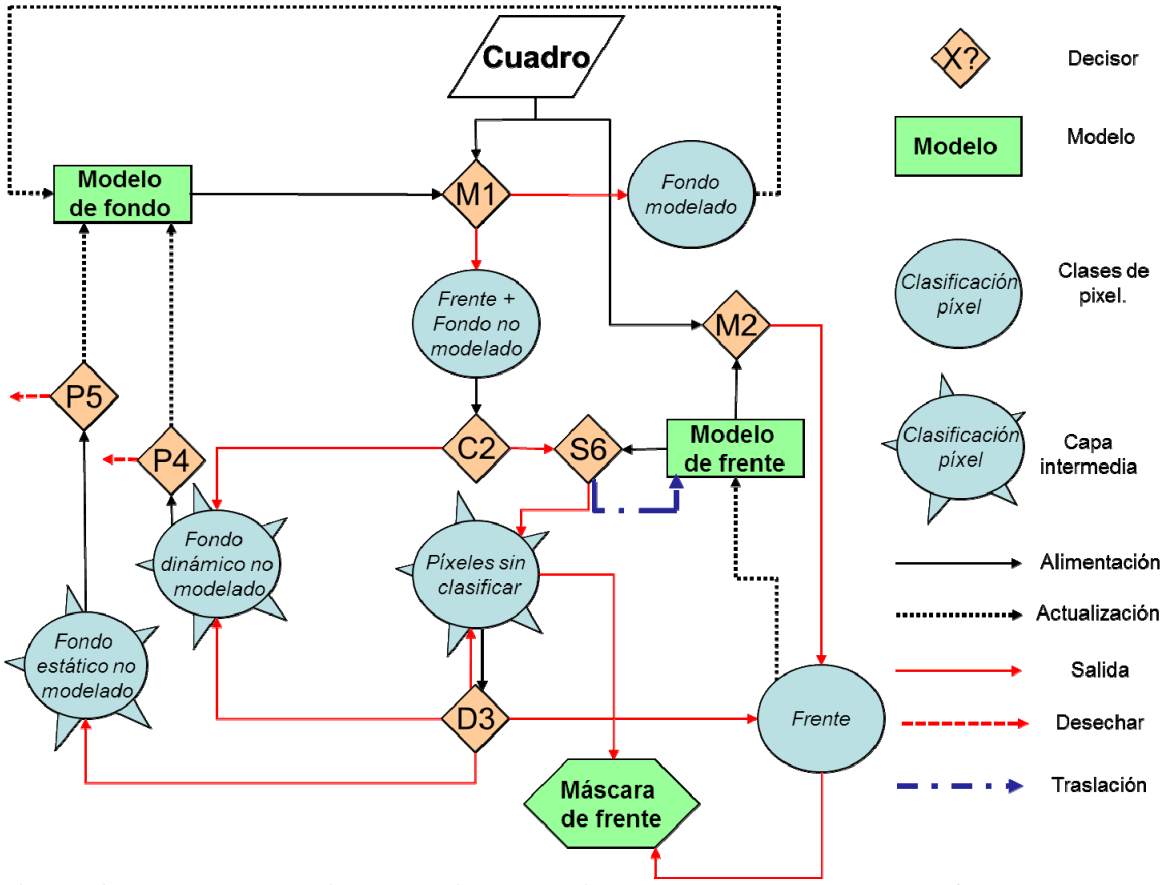


Figura 6-5: Esquema del sistema mejorado a nivel de blob con modelado de frente.

6.4.1.1 **Descripción de la arquitectura del sistema.**

En esta mejora se introduce un modelado de frente que permite detectar píxeles de frente mal clasificados debido al problema de camuflaje entre los píxeles de *frente* y el modelo de fondo. Sólo se describen los nuevos módulos que se han introducido en el sistema. El resto ya se definieron en 6.3.3.2.

Módulo S6

Utiliza un filtro de *Kalman* [42] que contiene los blobs incluidos en el modelo de frente y los nuevos blobs detectados en el nuevo cuadro de entrada que salen del módulo C2.

Los blobs que no se han detectado en el modelo de frente se clasifican como *píxeles sin clasificar*. El resto de blobs, los detectados como nuevas instancias de los contenidos en

el modelo de frente (marcados con la misma *BlobId*), pasan al modelo de frente para realizar la traslación del modelo.

Se destaca que este filtro de *Kalman* ha sido modificado para que no incluya los nuevos blobs detectados en memoria. Sólo identifica los blobs contenidos en el modelo de frente. Es decir, sólo se actualizan en memoria los blobs que existen en el modelo de frente, los blobs procedentes de C2 no se almacenan en memoria, puesto que su clasificación como *frente* no ha sido decidida aún.

Modelo de Frente

En el módulo del modelo de frente se modelan todos los píxeles pertenecientes a los blobs que se han clasificados como *frente* en el módulo decisor D3. Los blobs contenidos en el modelo son los que alimentan a S6.

El modelo de frente sólo dispone de una capa (en comparación con el de fondo que dispone de *k* capas). Esta capa de frente se inicializa y se actualiza igual que las capas del modelo de fondo (proceso descrito en el capítulo 4 y Anexo A).

La diferencia fundamental con el modelo de fondo es que mientras en el modelo de fondo, las apariencias del pixel se modelan en una posición fija, en el modelo de frente la posición de un modo va desplazándose en cada cuadro. Es decir, en cada cuadro se hace una traslación de los blobs incluidos en el modelo de frente.

Traslación del modelo de frente: con la correspondencia entre las nuevas instancias de los blobs modelados y los que se encontraban en el modelo de frente se evalúa el desplazamiento experimentado por el blob entre la nueva instancia y la última presente en el modelo.

El vector desplazamiento tiene como módulo el movimiento entre las dos instancias del blob (definido en la ecuación *eq 6.3*), es decir, su dirección y sentido se calculan mediante la ecuación:

$$\vec{\Delta} = (x_{t\text{-}n}^{PC} - x_t^{PC}, y_{t\text{-}n}^{PC} - y_t^{PC}) \qquad \textit{eq 6.4}$$

La traslación se realiza aplicando el vector de desplazamiento a los modos pertenecientes a la instancia del blob en el modelo de frente (es decir, a la instancia que existía previamente).

Una vez que se ha completado la traslación del modelo de frente se compara el cuadro de entrada con el modelo de frente mediante el módulo M2 para clasificar nuevos píxeles de frente que se han podido perder por camuflaje con el fondo.

Módulo M2

Es un módulo de funcionamiento idéntico a M1 pero con distinta alimentación que determina si los píxeles del cuadro de entrada han sido previamente modelados en el modelo de frente. Por ello, la diferencia fundamental (obvia y única) entre los dos módulos es que el primero utiliza para la comparación interna a la capa, las capas que modelan los píxeles clasificados como *fondo modelado* (modelo de fondo) y el segundo la que modela los clasificados como *frente* (modelo de frente).

6.4.1.2 **Limitaciones del sistema final.**

Con esta mejora se van a obtener mejores resultados en secuencias donde el *frente* tiene un aspecto parecido al *fondo modelado* y donde no se podía realizar una segmentación eficaz utilizando solamente modelado de fondo para la segmentación.

Puesto que éste es el último sistema implementado, es necesario indicar las limitaciones existentes que aún, incluso teóricamente, presenta este último sistema.

Existen dos limitaciones principales. La primera concierne a la dependencia del sistema a la naturaleza de las secuencias de entrada ya que aunque se ha normalizado el movimiento con la información de la resolución de la secuencia, este es también dependiente del tamaño del objeto de frente.

Por otro lado el sistema es capaz de recuperar píxeles camuflados por aspecto parecido al *fondo modelado* pero siempre que en el proceso de evaluación de pertenencia al *fondo modelado* se detecte si no una instancia del blob completo, al menos un sub-blob lo suficientemente similar al modelado como para que el algoritmo de seguimiento pueda

identificarlo como una nueva instancia del blob en el modelo de frente. Si la nueva instancia del blob no cumple esta premisa, el módulo S6 no es capaz de realizar el seguimiento y por lo tanto la traslación del modelo de frente no es posible. Situación que deriva en la imposibilidad de recuperar los píxeles de frente camuflados. Todo ello convierte al modelo de frente en dependiente del modelo de fondo.

6.5 Resumen de las mejoras y sistemas implementados para su evaluación.

En esta sección se hace un resumen de las mejoras implementadas y de los problemas que se han intentado resolver con cada una de ellas. Se enumeran en una tabla los sistemas que se han implementado para conseguir las mejoras objetivo.

Mediante la Tabla 6-2 puede intuirse la evolución de los resultados obtenidos con cada una de mejoras. Estos resultados se evaluarán cualitativa y cuantitativamente en el capítulo 7. La tabla incluye las características de cada mejora (ordenadas por sistema), los objetivos buscados con cada una de ellas y los problemas asociados que motivan el diseño del siguiente sistema.

Nombre	Características	Objetivos	Problemas
Bayesiano Básico (Sistema 1)	Inspirado en [1]. Bayesiano Básico con covarianza diagonal	Modelar las distintas apariencias del pixel en capas independientes mediante aprendizaje Bayesiano	Introducción de apariencias pertenecientes a objetos de frente en el modelo de fondo

Covarianza Completa (Sistema 2)	Bayesiano Básico con covarianza completa	Mejorar el cálculo de pertenencia a cada capa incluyendo información sobre la correlación temporal de las dimensiones de los datos de entrada.	Introducción de apariencias pertenecientes a objetos de frente en el modelo de fondo
Clasificación de pixel a bajo nivel (Sistema 3)	Propone una nueva clasificación de pixel que se realiza analizando características de bajo nivel	Evitar la introducción de modos pertenecientes a objetos de frente en el modelo de fondo.	No se modelan los *fondos estáticos no modelados* en el modelo de fondo con la rapidez que se modelaban en el sistema *Bayesiano Básico.*
Clasificación de pixel a nivel de blob (Sistema 4)	Sigue la clasificación de píxel anterior (*Sistema 3*) pero incluye características de blob para la clasificación	-Mejorar la clasificación de bajo nivel con las nuevas características. -Reducir el tiempo de inicialización del sistema y eliminar umbral de Confianza	Las nuevas características dependen de las secuencias de entrada.
Modelo de frente (Sistema 5)	Añade el modelado de frente al Si*stema 4*	Clasificar correctamente píxeles de *frente* mal clasificados por camuflaje con el *fondo modelado*.	El modelo de frente es dependiente del modelo de fondo.

Tabla 6-2. Resumen de las mejoras y sistemas implementados.

7 **Resultados.**

Los resultados de este proyecto fin de carrera se han obtenido mediante la evaluación de las máscaras de segmentación que se han obtenido por medio de los diferentes algoritmos desarrollados.

Estas técnicas de evaluación se pueden clasificar en dos grandes grupos: técnicas subjetivas o cualitativas que dependen del criterio humano de evaluación y técnicas objetivas o cuantitativas, que suelen utilizar una segmentación de referencia, bien anotada a mano o bien mediante grabación en estudios de croma [44], denominada *ground-truth*, para compararla con la segmentación obtenida por el algoritmo de segmentación implementado. Estas medidas permiten evaluar la calidad de la segmentación.

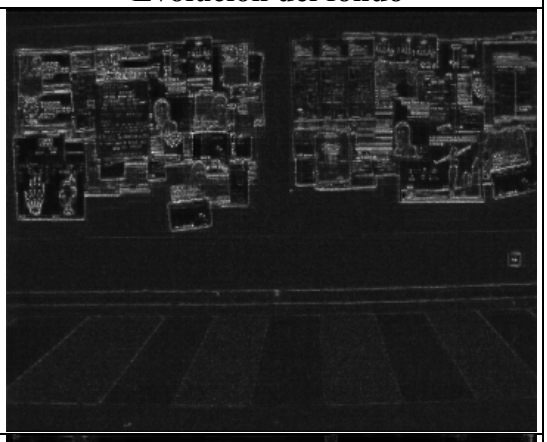
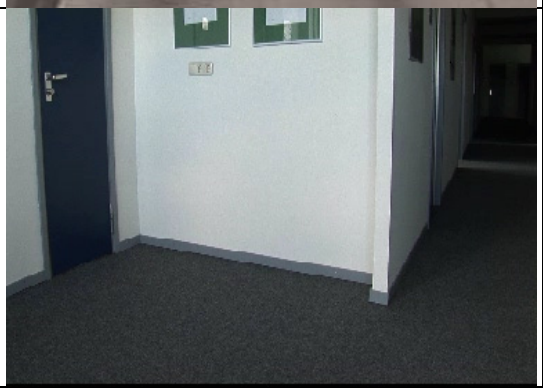
Para conseguir los resultados y poder hacer una evaluación de los algoritmos implementados en este proyecto se ha trabajado con un conjunto de secuencias de video de diferentes características que constituyen el grupo de videos de prueba. Este conjunto de vídeos de prueba se describe en el siguiente apartado.

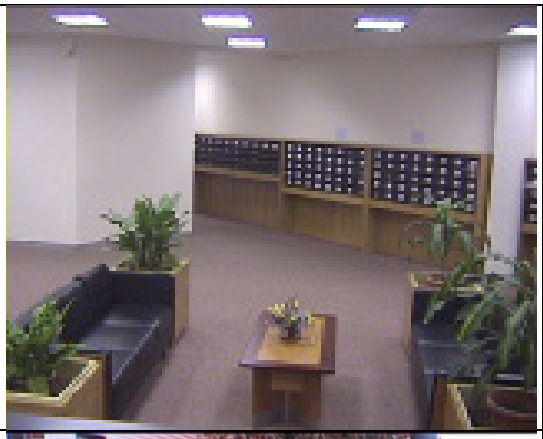
7.1 **Descripción de las secuencias de prueba.**

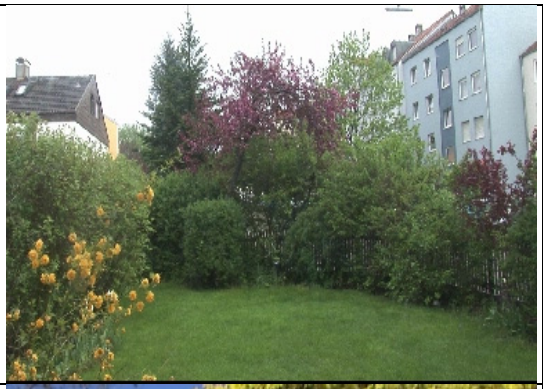
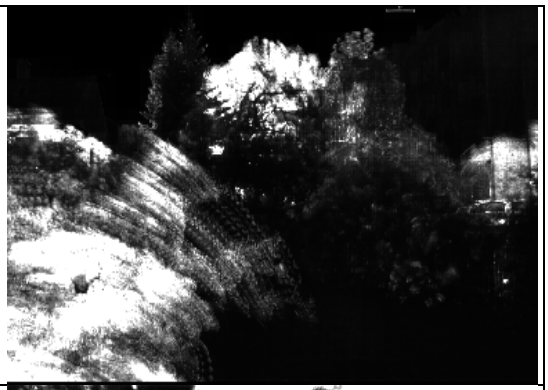
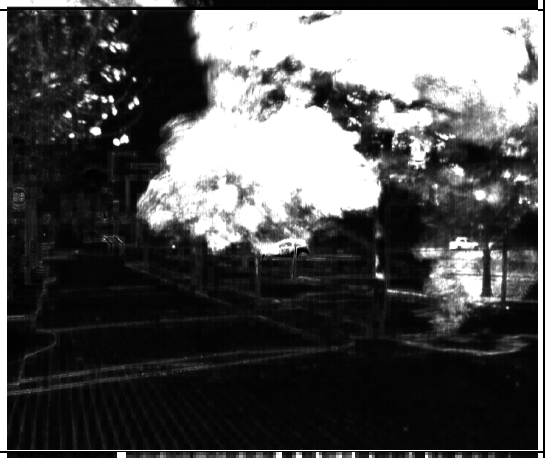
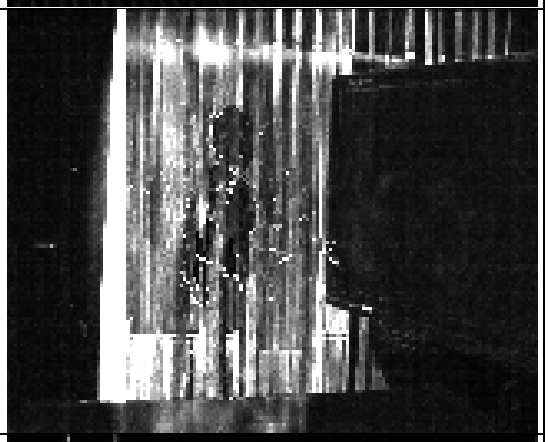
Las secuencias de video utilizadas están clasificadas por su complejidad, es decir, por el grado de dificultad que presenta para un algoritmo de segmentación, analizar dichas secuencias. Este grado de dificultad depende de una serie de propiedades denominadas '*factores críticos*' entre las que se encuentran las características específicas de los objetos en movimiento, el tipo de fondo o el movimiento de la cámara (aunque en nuestro caso, se ha trabajado con secuencias de video sin movimiento de cámara o de cámara estática).

En la Tabla 7-1 se muestra en la primera columna el identificador asignado a cada secuencia para diferenciarlas '*ID*'. En la segunda columna se muestra un cuadro del vídeo que permite ver las características del mismo. Por último y para ofrecer una descripción intuitiva de la variabilidad del fondo de escena y su complejidad, se muestra en la tercera columna una imagen donde se recoge la variación media (más alta cuanto más blanca) de

cada pixel a lo largo del vídeo o simplemente la '*Evolución del fondo*', esta evolución ha sido normalizada para facilitar su visualización.

ID	Cuadro	Evolución del fondo
S1		
S2		
S3		
S4		

S5		
S6		
S7		
S8		

S9		
S10		
S11		
S12		

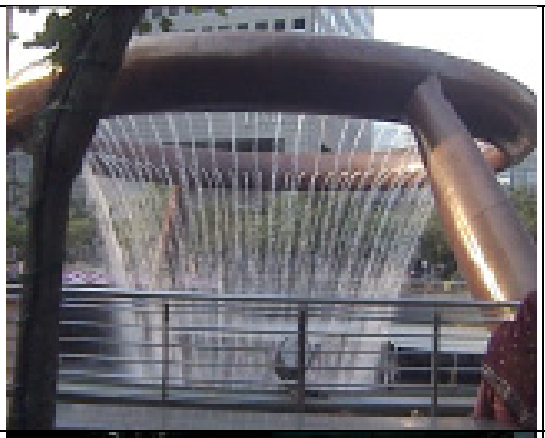
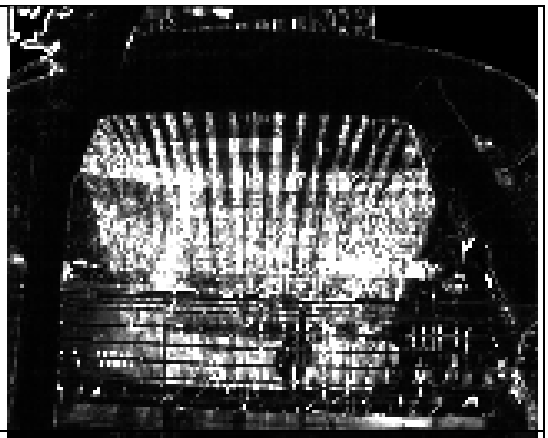
S13		
S14		

Tabla 7-1. Secuencias de prueba.

En la Tabla 7-2 se enumeran los diferentes videos utilizados, indicando su *ID*, nombre, y DataSet asociado donde se suministran las máscaras de *ground-truth*. Adicionalmente, se describen una serie de factores críticos para cada una de las secuencias de los vídeos de prueba. En base a estos factores, se realiza una clasificación de la complejidad de la secuencia: baja, media y alta:

ID	Nombre vídeo	Modalidad del fondo	Sombras	Cambios de iluminación	Complejidad	Data Set
S1	**Baile Unimodal**	Unimodal.	No	No	Baja	[44]
S2	**Laboratory**	Unimodal.	Si	Si	Alta	[31]
S3	**Highway**	Unimodal.	Si	No	Media	[31]
S4	**Pasillo**	Unimodal.	No	Si	Media	[45]
S5	**Lobby**	Unimodal.	Si	Si	Alta	[46]
S6	**Ingravidez**	Multimodal alto.	No	No	Alta	[44]
S7	**Silla**	Multimodal medio.	No	No	Alta	[44]
S8	**Hambre**	Multimodal bajo.	No	No	Baja	[44]
S9	**Jardín**	Multimodal alto.	No	No	Alta	[45]
S10	**Baile Multimodal**	Multimodal alto.	No	No	Alta	[44]
S11	**Curtain**	Multimodal medio.	Si	Si	Alta	[46]
S12	**Water Surface**	Multimodal alto.	No	No	Media	[46]
S13	**Fountain**	Multimodal alto.	Si	No	Media	[46]
S14	**Escalator**	Multimodal alto.	Si	Si	Alta	[46]

Tabla 7-2. Características de las secuencias de prueba.

A continuación se muestra el diseño de pruebas utilizado para evaluar los algoritmos implementados. Para ello se utiliza el conjunto de secuencias de prueba que se acaba de presentar. El diseño de las pruebas experimentales se ha realizado con el objetivo de obtener resultados comparativos de los distintos sistemas implementados para observar la influencia de las distintas mejoras en la calidad final de los resultados. Se ha realizado una

comparativa de coste computacional entre los distintos sistemas pero es en la calidad de la segmentación donde se ha trabajado más en este proyecto.

7.2 .Configuración inicial del sistema.

Los parámetros y umbrales utilizados en este proyecto son internos del sistema y no se cambian en función de la secuencia a segmentar. La inicialización de parámetros para las diferentes pruebas utilizada ha sido la siguiente:

En cuanto al número máximo de capas (el parámetro *k* configurable) se han realizado pruebas para *k=3* y *k=5*. Cada vez que se crea una nueva capa en el sistema implementado se inicializan las variables propias de la capa según se describe en el Anexo A.

Para la inicialización de los parámetros descritos en el capítulo 4 necesarios para modelar la distancia de Mahalanobis en cada capa (μ_t y σ_t), se han utilizado diferentes esquemas en función de la información disponible a lo largo de la ejecución del algoritmo.

En los sistemas *Bayesiano Básico (Sistema 1)*, *Bayesiano Básico* con matriz de covarianza completa *(Sistema 2)* la inicialización utilizada es la siguiente:

$$\mu_0 = 0$$
$$\sigma_0 = 2 \qquad \textit{eq 7.1}$$

Para la clasificación de pixel utilizando información extraída a nivel de blob *(Sistema 4)*, la inicialización es diferente puesto que se aprovecha la distinción entre los píxeles de *fondo estático no modelado* y los píxeles de *fondo dinámico no modelado* definidos en 6.3.1.

Para los píxeles clasificados como *fondo estático no modelado* los umbrales permitidos son mayores al considerarse que sus modos más fiables que los de los píxeles clasificados como *fondo dinámico no modelado*, por ser éstos últimos más propensos a ser erróneamente clasificados como *frente* o de apariencias similares al *fondo modelado*.

La inicialización para los píxeles clasificados como *fondo estático no modelado* es:

$$\mu_0 = 0 \qquad \sigma_0 = 3 \qquad \textit{eq 7.2}$$

Para los clasificados como *fondo dinámico no modelado*:

$$\mu_0 = 0 \qquad \sigma_0 = 1 \qquad \textit{eq 7.3}$$

Para el sistema que utiliza información a nivel de píxel (*Sistema 3*) podría utilizarse la misma inicialización que para el *Sistema 4*, puesto que teóricamente se conseguiría la misma distinción entre los dos tipos de píxeles de fondo no modelados. En cambio, los resultados indican que esa distinción no se produce con suficiente fiabilidad en este sistema, por lo que decidimos inicializarlos mediante los parámetros descritos en la *eq 7.1*.

El *Sistema 5*, al tratarse de una evolución del *Sistema 4* donde sólo se ha añadido el modelo de frente, se inicializará de manera equivalente.

Por otro lado, para establecer el proceso de umbralización automática de la distancia (ver sección 4.3.2) se ha utilizado un factor de aprendizaje α configurable. En lugar de fijar un valor de α estático, o hacerlo dependiente del tiempo desde el comienzo de la ejecución, en todos los sistemas desarrollados en este proyecto final de carrera se ha utilizado un factor de aprendizaje que depende de la confianza de cada capa.

El objetivo, es conseguir que el sistema se adapte automáticamente a las características de la secuencia de entrada. Los parámetros fijados son empíricos, pero responden a las intuiciones teóricas descritas a continuación.

Se ha detectado que cuando los modos tienen confianzas altas, aumenta la probabilidad de que exista sobre-aprendizaje (es decir, que el sistema se ajuste demasiado al valor de la moda y permita pocas variaciones). Por otro lado, cuando tienen confianzas bajas (modos poco fiables) el sistema puede ser demasiado laxo en el modelado de la distancia, otorgando condiciones de pertenencia a algunas muestras que realmente se alejan

de la moda. Para intentar evitar estos problemas o al menos minimizarlos, se ha utilizado la siguiente configuración para los valores de α :

$$\alpha = 0.01 \xrightarrow{Si} C < \frac{U_C}{2}$$
$$\alpha \propto 1/C \;, \xrightarrow{Si} \frac{U_C}{2} < C < U_C \qquad \textit{eq 7.4}$$
$$\alpha = 0.001 \xrightarrow{Si} C > U_C$$

Donde α es o bien fija o bien proporcional a la confianza ($\alpha = \frac{1}{\kappa}$, donde κ se define en el Anexo A).

Con el mismo objetivo, se establecen parámetros de acotación sobre la desviación típica que evitan un estrechamiento excesivo de la *fdp* Gaussiana que modela la evolución, en cada capa, de la distancia de Mahalanobis.

$$\sigma_t(x,y) > 1 \xrightarrow{Si} 0,5 < C < 1$$
$$\sigma_t(x,y) > 2 \xrightarrow{Si} C > 1 \qquad \textit{eq 7.5}$$

Es decir, la desviación típica no puede bajar de *1* si la confianza está comprendida entre *0,5* y *1*, ni bajar de *2* si ésta es superior a *1*.

Finalmente, la siguiente Tabla resume los umbrales empíricos aplicados sobre las características analizadas a nivel de blob utilizadas para realizar la clasificación de pixel en la aproximación descrita en la sección 6.3.3.

Clase	Solape	Movimiento
fondo modelado	Irrelevante	Irrelevante
fondo estático no modelado	>0.98	< 0.05
frente	> 0.6	> 0.95
	> 0.8	>0.7 & < 0.95
fondo dinámico no modelado	<0.6	< 0.05
	<0.4	> 0.05 & < 0.2

Tabla 7-3. Clasificación de pixel en función de la variabilidad de la confianza.

7.3 **Análisis comparativo de los sistemas implementados.**

A lo largo de esta sección, se exponen los resultados obtenidos en las ejecuciones de los sistemas implementados sobre las secuencias de prueba.

En el primer apartado, se comparan los tiempos de ejecución de los algoritmos implementados. A continuación se analiza cuantitativamente la calidad de los diferentes métodos de segmentación desarrollados en este proyecto para evaluar la aportación de cada una de las mejoras desarrolladas. Para el análisis cuantitativo se han utilizado métricas comunes en problemas de clasificación, que son explicadas en el apartado 7.3.2. Finalmente, para cada video se presentan los resultados gráficos (la máscara de segmentación) de un cuadro considerado 'conflictivo' del video. Este cuadro intenta ofrecer una descripción cualitativa del comportamiento de cada uno de los sistemas.

7.3.1 **Coste computacional de cada uno de los sistemas.**

En primer lugar, se ha realizado una comparativa en coste computacional (tiempo de ejecución) de cada algoritmo implementado con cada secuencia de prueba. El tiempo se ha medido mediante ejecución sobre un procesador Pentium IV @ 2.8 Ghz con 2GB de RAM. Los tiempos calculados se expresan en segundos de procesamiento por cuadro.

Primero se ofrecen tiempos medios por cuadro por algoritmo y por secuencia (Tabla 7-4. Tiempos de procesamiento utilizando modelado con k=3 para el modelo de fondo. y Tabla 7-5), después, estos tiempos se utilizan para calcular el tiempo medio por algoritmo (Figura 7-1).

TIEMPOS DE PROCESAMIENTO (s/cuadro) para k=3 capas					
ID	Bayesiano básico	Covarianza completa	Clasificación con información de bajo nivel	Clasificación con información de alto nivel	Modelado de frente
S1	1,568	1,289	1,383	0,985	1,089
S2	0,981	0,883	1,017	0,684	0,731
S3	0,886	0,882	1,056	0,704	0,790
S4	1,429	1,112	1,334	0,864	0,930
S5	0,168	0,179	0,206	0,142	0,154
S6	1,656	1,674	1,846	1,112	1,220
S7	1,422	1,384	1,603	1,028	1,142
S8	1,235	1,232	1,436	0,965	1,076
S9	1,754	1,522	1,819	1,071	1,170
S10	1,434	1,428	1,536	1,132	1,283
S11	0,186	0,194	0,209	0,148	0,160
S12	0,207	0,171	0,192	0,142	0,152
S13	0,192	0,180	0,194	0,145	0,155
S14	0,213	0,217	0,232	0,168	0,177

Tabla 7-4. Tiempos de procesamiento utilizando modelado con k=3 para el modelo de fondo.

TIEMPOS DE PROCESAMIENTO (s/cuadro) para k=5 capas					
ID	Bayesiano básico	Covarianza completa	Clasificación con información de bajo nivel	Clasificación con información de alto nivel	Modelado de frente
S1	2,598	2,447	2,785	1,373	1,483
S2	1,691	1,701	1,925	0,964	1,010
S3	1,598	1,707	1,969	0,981	1,065
S4	2,291	2,119	2,449	1,204	1,271
S5	0,305	0,339	0,386	0,203	0,214
S6	2,703	2,884	3,198	1,512	1,615
S7	2,440	2,556	2,895	1,420	1,535
S8	2,235	2,410	2,708	1,356	1,473
S9	2,660	2,570	2,758	1,412	1,542
S10	2,462	2,637	2,698	1,526	1,671
S11	0,335	0,364	0,371	0,213	0,225
S12	0,346	0,342	0,336	0,200	0,211
S13	0,334	0,342	0,348	0,202	0,214
S14	0,390	0,417	0,416	0,228	0,240

Tabla 7-5. Tiempos de procesamiento utilizando modelado con k=5 máximas para el fondo.

Para comparar visualmente los tiempos de procesamiento de cada cuadro por cada método, se ha realizado un diagrama de barras con las medias de los tiempos obtenidos por método.

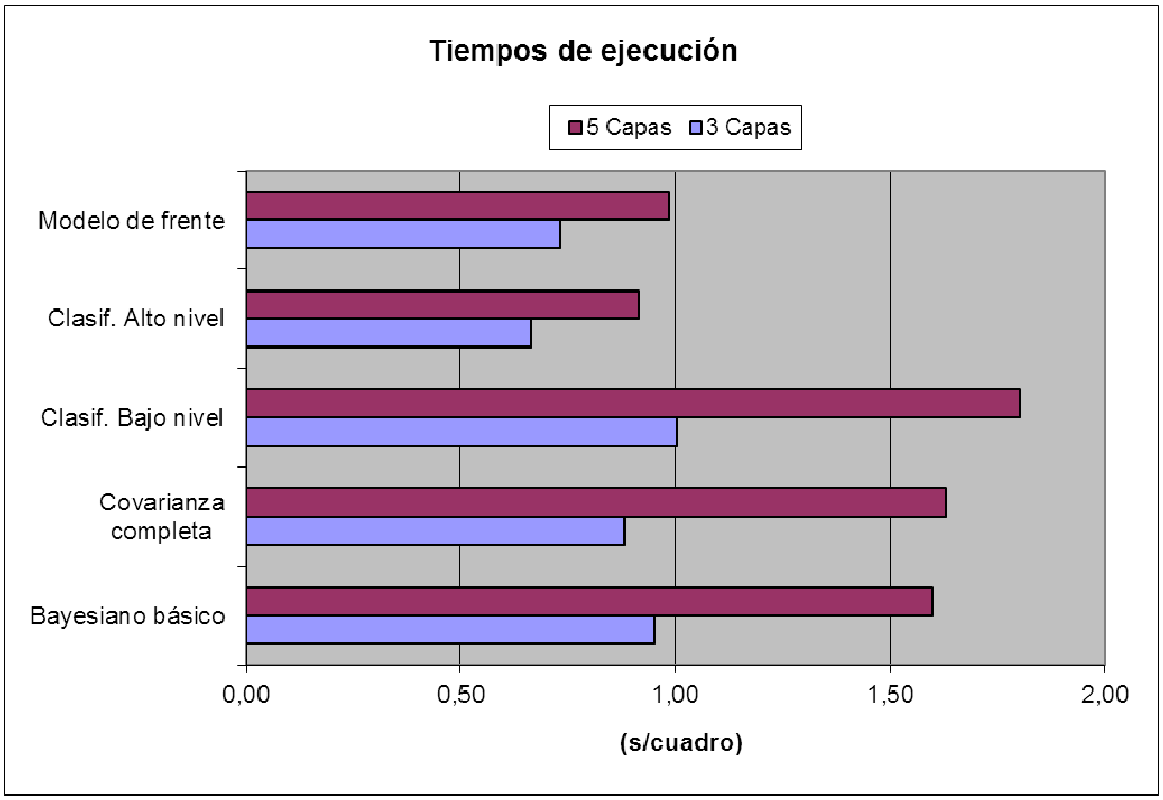


Figura 7-1: Tiempos medios de procesamiento de los algoritmos desarrollados.

Como se observa en las tablas y en el diagrama, el aumentar el número de capas utilizadas para el modelado del fondo aumenta significativamente el coste computacional del sistema. Esto se observa en todos los algoritmos, desde el sistema básico hasta el sistema con el modelado de frente.

Hay que destacar que estas medidas de coste computacional no son del todo fiables y hay que verlas en el contexto del proyecto. Como se observa en el diagrama de barras Figura 7-1, el sistema que incorpora la clasificación de pixel utilizando información a nivel de blob es el que tiene menor coste computacional de todos los algoritmos implementados y el *Bayesiano Básico* con matriz de covarianza diagonal es el que tiene mayor coste computacional. Esto es debido a varias razones:

- El código implementado se ha ido optimizando a lo largo del proyecto y por eso las últimas mejoras implementadas tienen menor coste computacional.

- Además hay que destacar que el código se desarrolló pensando en utilizar la matriz de covarianza completa. Por ello el algoritmo básico con covarianza diagonal tiene un coste computacional similar al algoritmo que incorpora la covarianza completa. Si se

hiciese una implementación considerando la covarianza diagonal, el coste computacional debería ser inferior.

En este trabajo fin de carrera se ha puesto mayor dedicación en ver como mejoraban los resultados en la detección de frente con cada una de las mejoras que en controlar el coste computacional que demandaban.

7.3.2 **Métricas utilizadas.**

Las métricas que se han utilizado para evaluar los métodos de segmentación desarrollados en este proyecto se basan en la comparación de las máscaras binarias del frente generadas por cada método y el *ground-truth* disponible para cada secuencia.

La comparación se realiza a nivel de pixel utilizando las medidas que se muestran a continuación.

- Verdaderos positivos (***TP***): es el número de detecciones correctas de valor uno (frente).

- Verdaderos negativos (***TN***): es el número de detecciones correctas de valor cero (fondo).

- Falsos positivos (***FP***): es el número de detecciones incorrectas de valor uno.

- Falsos negativos (***FN***): es el número de detecciones incorrectas de valor cero.

En la siguiente tabla se muestran estas medidas y sus relaciones con la máscara de *ground truth* y con la máscara obtenida a la salida del algoritmo de segmentación.

		Ground truth		
		Positivos reales	Negativos reales	
Algoritmo de segmentación	Positivos detectados	**TP**	**FP**	TP+FP
	Negativos detectados	**FN**	**TN**	FN+TN
		TP+FN	FP+TN	

Tabla 7-6. Parámetros estadísticos.

Las medidas anteriores suelen combinarse y dar lugar a otras muy usuales en el estado del arte de segmentación de objetos. Estas medidas son las siguientes:

- **Precisión:** se define como el número total de píxeles correctos de un tipo con respecto al total de los píxeles de este tipo detectados por nuestro algoritmo de segmentación. Las máscaras que se obtienen de la segmentación son binarias presentando píxeles con valor cero y valor uno. El objetivo a la hora de evaluar los algoritmos no es únicamente la correcta detección del objeto de frente (*máscara=1*) sino también, que los píxeles en ausencia de movimiento sean correctamente detectados como fondo (*máscara=0*), esto supone el cálculo de Precisión en los dos grupos:

$$\textbf{\textit{Precisión}}\left(máscara = 0\right) = P0 = \frac{TN}{TN + FN} \qquad eq\ 7.6$$

$$\textbf{\textit{Precisión}}\left(máscara = 1\right) = P1 = \frac{TP}{TP + FP} \qquad eq\ 7.7$$

- **Recall**: es el número de píxeles correctos detectados de un tipo con respecto al total real de ese tipo marcado por el '*ground truth'*. Por el mismo motivo que en el caso anterior, existen dos grupos de valores en las máscaras obtenidas a la salida de los sistemas implementados y por ello, se ha calculado un Recall para cada conjunto:

$$\textbf{\textit{Recall}}\ \left(máscara = 0\right) = R0 = \frac{TN}{TN + FP} \qquad eq\ 7.8$$

$$\textbf{\textit{Recall}}\left(máscara = 1\right) = R1 = \frac{TP}{TP + FN} \qquad eq\ 7.9$$

- **F-score:** Una forma de combinar la correcta detección del algoritmo con respecto a las detecciones realizadas por el mismo (Precisión) con la correcta detección del algoritmo con la realidad (Recall) es a través del F-score, que concede a ambas medidas igual peso. El F-score también ha sido calculado para cada conjunto de valores de la máscara binaria resultado de cada sistema:

$$\textbf{\textit{F - score}}\left(máscara = 0\right) = FS0 = \frac{2 \cdot P0 \cdot R0}{P0 + R0} \qquad eq\ 7.10$$

$$\textbf{\textit{F - score}}\left(máscara = 1\right) = FS1 = \frac{2 \cdot P1 \cdot R1}{P1 + R1} \qquad eq\ 7.11$$

Para lograr el objetivo de evaluar y encontrar los parámetros óptimos de los algoritmos se han ponderado de manera equivalente la detección de frente/fondo en la máscara binaria fusionando las medidas anteriores a través de una suma [2] (si bien se podía haber utilizado cualquier otra función). La selección del parámetro óptimo se realiza eligiendo la suma máxima:

$$Parámetro\_óptimo = FS0 + FS1 \qquad eq\ 7.12$$

7.3.3 **Calidad de la segmentación.**

Para el análisis comparativo se han obtenido en cada caso, tablas y diagramas de barras de las medidas de calidad (descritas en el apartado anterior 7.3.2): Precisión en cero (**P0**), Precisión en uno (**P1**), Recall en cero (**R0**), Recall en uno (**R1**), F-score en cero (**FS0**) y F-score en uno (**FS1**). Todas estas medidas son las calculadas para las distintas secuencias de prueba 7.1 con cada uno de los sistemas desarrollados en este proyecto fin de carrera.

Estos resultados comparativos se han desarrollado mediante comparativa de los resultados de los distintos algoritmos para cada una de las secuencias de prueba. Para cada secuencia se muestran dos tablas con los resultados obtenidos por los distintos algoritmos. En la primera se fija *k=3* como número máximo de capas para almacenar el modelo de fondo y en la segunda *k=5*.

Posteriormente se muestran cuatro gráficas para comparar visualmente los resultados obtenidos: una para la Precisión, otra para el Recall, otra para el F-score y una última con el parámetro óptimo (**FS0+ FS1**) [2]. Estas gráficas solo se muestran para la configuración con *k=3* capas ya que como se observa en las tablas prácticamente no hay diferencias en la detección. En las tablas se marca en rojo los mejores resultados y en ***negrita*** el sistema que se comporta mejor globalmente.

Para complementar el análisis comparativo se ha añadido una comparación subjetiva para cada una de las secuencias de prueba mostrando segmentaciones obtenidas por cada uno de los sistemas implementados donde se puede observar los principales problemas de cada una de las técnicas.

Finalmente, es importante destacar que para todas las secuencias analizadas, se han calculado los estadísticos utilizando las máscaras de segmentación a partir del cuadro *50*, con el objetivo de no analizar el periodo de inicialización necesario (para pasar de *fondo no fiable* a *fondo fiable*) en los *Sistemas 1,2 y 3*. En caso contrario, los estadísticos de los *Sistemas 4* y *5*, que no requieren de este tiempo de inicialización presentarían magnitudes similares, mientras que *1,2 y* 3 resultados mucho peores.

➢ S1: Baile Unimodal.

Características de las secuencias S1: interior, unimodal y de baja complejidad.

Resultados cuantitativos.

Presentados por medio de tablas y diagramas de barras.

SECUENCIA S1: BaileUni 3 Capas						
Método de Sustracción de fondo.	**P0 %**	**P1 %**	**R0 %**	**R1 %**	**FS0 %**	**FS1 %**
Bayesiano básico con covarianza diagonal. (*Bayesiano Básico*) *Sistema 1*	97,93	96,00	99,64	80,55	98,78	87,60
Bayesiano básico con covarianza completa. (*Covarianza completa*) *Sistema 2*	97,58	96,93	99,74	77,12	98,64	85,90
Mejora utilizando clasificación de pixel de bajo nivel. (*Clasificación bajo nivel*) *Sistema 3*	99,87	97,41	99,72	98,80	99,79	98,10
Mejora utilizando clasificación de pixel de alto nivel. (*Clasificación a nivel de blob*) *Sistema 4*	99,87	97,46	99,72	98,77	99,79	98,11
Clasificación a nivel de blob y modelado de frente. (*Modelado de frente*) *Sistema 5*	99,93	97,36	99,71	99,34	99,82	98,34

Tabla 7-7. Resultados comparativos sobre la secuencia S1 con modelo de fondo de 3 capas.

SECUENCIA S1: BaileUni 5 Capas						
Método de Sustracción de fondo.	**P0 %**	**P1 %**	**R0 %**	**R1 %**	**FS0 %**	**FS1 %**
Bayesiano Básico Sistema 1	97,83	95,97	99,64	79,61	98,73	87,03
Covarianza completa Sistema 2	97,39	96,87	99,74	75,31	98,55	84,74
Clasificación bajo nivel Sistema 3	99,87	97,41	99,72	98,80	99,79	98,10
Clasificación nivel de blob Sistema 4	99,87	97,46	99,72	98,77	99,79	98,11
Modelado de frente Sistema 5	99,93	97,36	99,71	99,34	99,82	98,34

Tabla 7-8. Resultados comparativos sobre la secuencia S1 con modelo de fondo de 5 capas.

A continuación se muestran los resultados por medio de diagramas de barras de la **Tabla 7-7**.

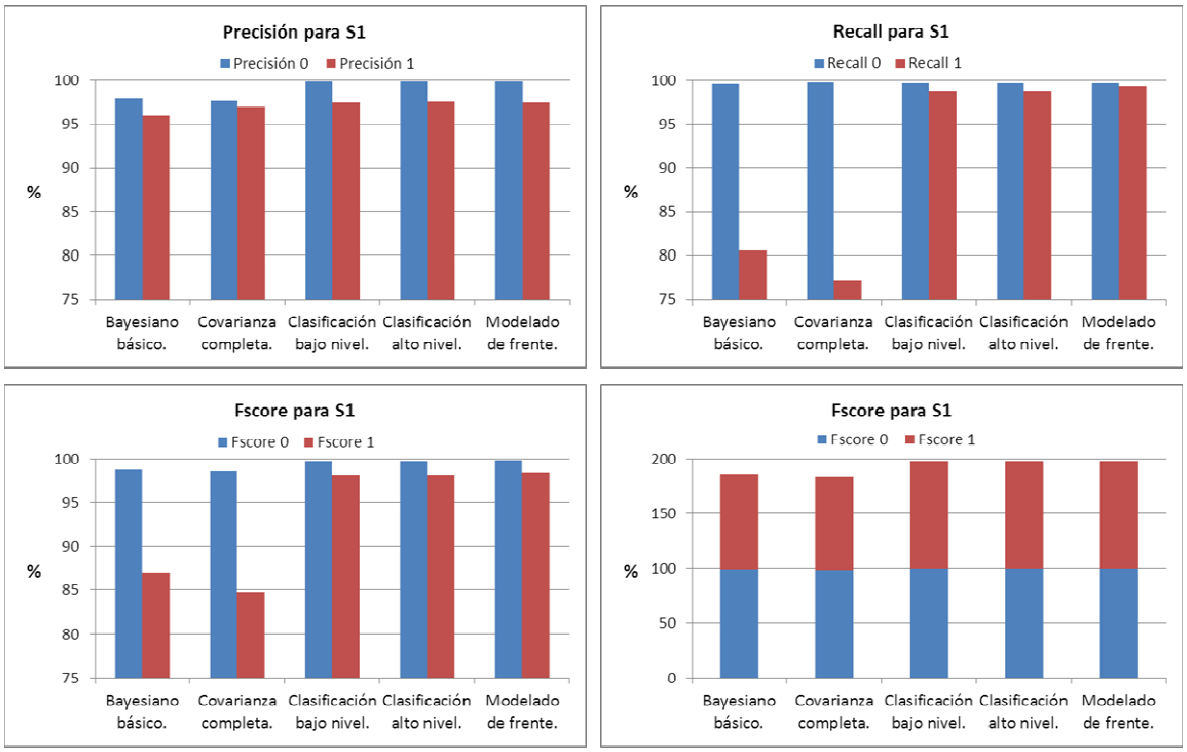


Figura 7-2: Diagramas de barras de los estadísticos de S1.

Resultados cualitativos.

Presentados para el cuadro 594 de la secuencia S1:

Cuadro			**Fondo**
GT			**Sistema1**
Sistema2			**Sistema3**
Sistema4			**Sistema5**

Tabla 7-9. Resultados cualitativos sobre la secuencia S1.

Discusión.

Como se observa en la **Tabla 7-7** y **Tabla 7-8**, los resultados obtenidos para esta secuencia de prueba presentan altos valores de Precisión y Recall, para todos los sistemas desarrollados.

En las tablas se observa que la detección del fondo está por encima del 98% en todos los sistemas y que las diferencias entre un sistema y otro se aprecian mejor en la detección del frente. Los sistemas que peor lo detectan son los sistemas que utilizan la clasificación definida en el *Bayesiano Básico (Sistqamas 1 y 2)* ya que el frente se mueve por la misma zona durante un tiempo prolongado y por lo tanto se introduce en el modelo de fondo. Esta fue una de las razones que inspiró las mejoras basadas en clasificación de pixel que se realizan en los tres últimos sistemas. Se observa el correcto comportamiento de los sistemas que incluyen estas mejoras, en estos sistemas la detección del frente se sitúa por encima del 98%.

- **S2: Laboratory.**

Características de las secuencias S2: interior, unimodal, con cambios de fondo, con sombras y de complejidad alta.

Resultados cuantitativos.

Presentados por medio de tablas y diagramas de barras.

SECUENCIA S2: Laboratory 3 Capas						
Método de Sustracción de fondo.	**P0** **%**	**P1** **%**	**R0** **%**	**R1** **%**	**FS0** **%**	**FS1** **%**
Bayesiano Básico Sistema 1	99,83	41,44	96,31	94,01	98,04	57,53
Covarianza completa Sistema 2	99,72	48,71	97,37	90,03	98,53	63,22
Clasificación bajo nivel Sistema 3	99,86	33,52	94,76	95,11	97,24	49,57
Clasificación nivel de blob Sistema 4	99,85	34,45	94,98	95,02	97,36	50,57
Modelado de frente Sistema 5	99,85	47,26	97,07	94,61	98,44	63,03

Tabla 7-10. Resultados comparativos sobre la secuencia S2 con modelo de fondo de 3 capas.

SECUENCIA S2: Laboratory 5 Capas						
Método de Sustracción de fondo.	**P0** **%**	**P1** **%**	**R0** **%**	**R1** **%**	**FS0** **%**	**FS1** **%**
Bayesiano Básico Sistema 1	99,83	41,64	96,34	93,97	98,05	57,70
Covarianza completa Sistema 2	99,71	48,68	97,38	89,70	98,53	63,11
Clasificación bajo nivel Sistema 3	99,86	33,52	94,76	95,11	97,24	49,57
Clasificación nivel de blob Sistema 4	99,85	34,45	94,98	95,02	97,36	50,57
Modelado de frente Sistema 5	99,85	47,24	97,07	94,61	98,44	63,01

Tabla 7-11. Resultados comparativos sobre la secuencia S2 con modelo de fondo de 5 capas.

A continuación se muestran los resultados visuales por medio de diagramas de barras de la Tabla 7-10.

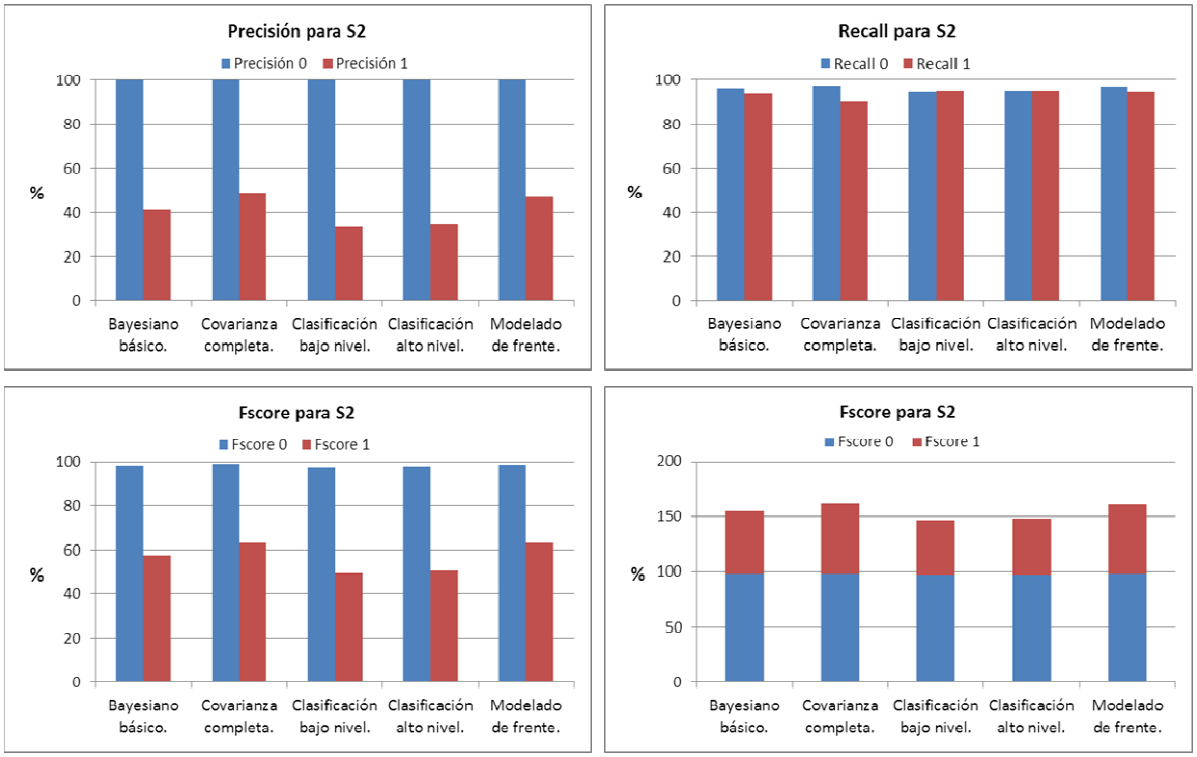


Figura 7-3: Diagramas de barras de los estadísticos de S2.

Resultados cualitativos.

Presentados para el cuadro 285 de la secuencia S2:

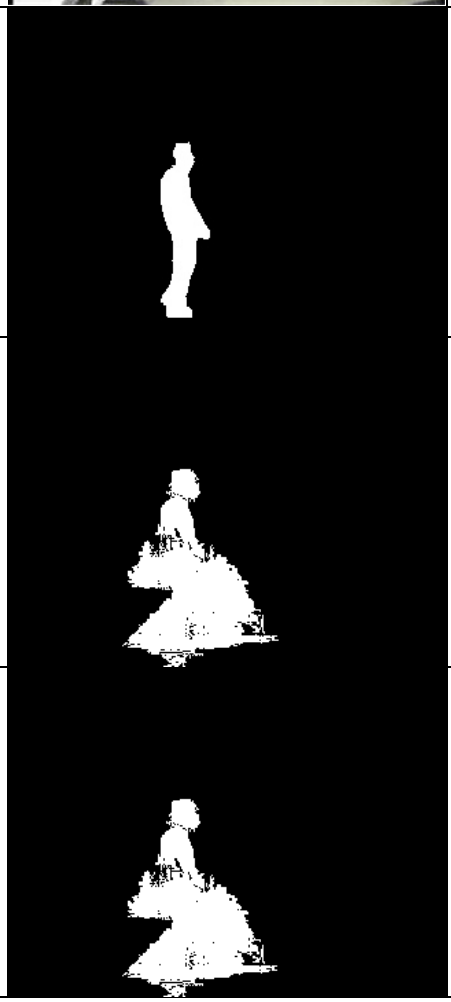
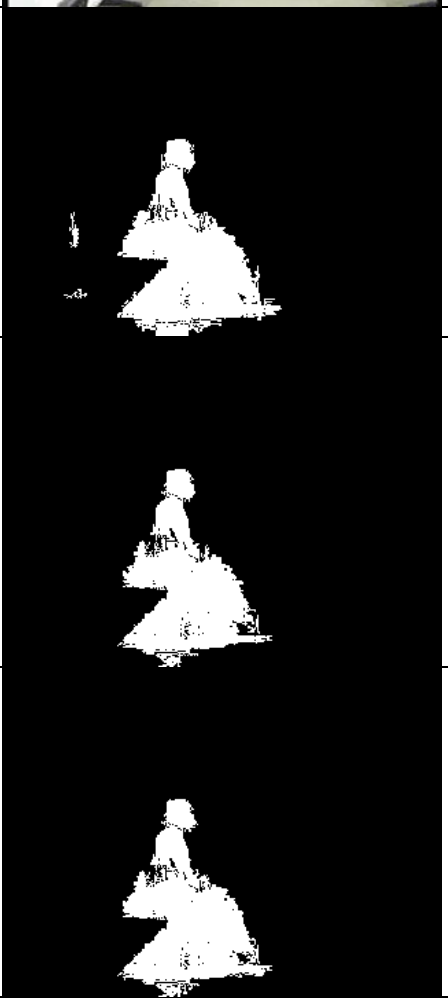
Cuadro			**Fondo**
GT			**Sistema1**
Sistema2			**Sistema3**
Sistema4			**Sistema5**

Tabla 7-12. Resultados cualitativos sobre la secuencia S2.

Discusión.

Como se observa en la Tabla 7-10 y Tabla 7-11, los resultados obtenidos se podrían mejorar sobre todo en la precisión de 1. En las tablas se observa que la detección del fondo está por encima del 97% en todos los sistemas pero que la detección de frente apenas supera el 60 % en los sistemas que aportan mejores resultados (*covarianza completa y modelo de frente*). Esto se debe a que el vídeo contiene sombras y que los sistemas

implementados en este proyecto no están pensados para segmentar vídeos que contengan sombras. El realizar un segmentador robusto a las sombras, bien mediante una etapa de post-procesado adicional, bien utilizando una nueva clase específica, queda como posible trabajo futuro para mejorar el trabajo realizado. En esta secuencia de prueba se observa como la mejora de introducir la matriz de covarianza completa para discriminar entre frente o fondo mejora considerablemente los resultados obtenidos con la matriz de covarianza diagonal.

- **S3: Highway.**

Características de las secuencias S3: exterior, unimodal, con sombras y de complejidad media.

Resultados cuantitativos.

Presentados por medio de tablas y diagramas de barras.

SECUENCIA S3: Highway 3 Capas						
Método de Sustracción de fondo.	**P0** **%**	**P1** **%**	**R0** **%**	**R1** **%**	**FS0** **%**	**FS1** **%**
Bayesiano Básico Sistema 1	99,31	41,50	85,45	94,59	91,86	57,69
Covarianza completa Sistema 2	98,87	47,68	89,14	90,70	93,76	62,50
Clasificación bajo nivel Sistema 3	97,93	32,68	81,05	84,32	88,70	47,11
Clasificación nivel de blob Sistema 4	96,54	45,43	90,80	70,21	93,58	55,17
Modelado de frente Sistema 5	96,69	49,83	92,19	71,10	94,39	58,60

Tabla 7-13. Resultados comparativos sobre la secuencia S3 con modelo de fondo de 3 capas.

SECUENCIA S3: Highway 5 Capas						
Método de Sustracción de fondo.	**P0** **%**	**P1** **%**	**R0** **%**	**R1** **%**	**FS0** **%**	**FS1** **%**
Bayesiano Básico Sistema 1	99,31	41,51	85,46	94,59	91,87	57,70
Covarianza completa Sistema 2	98,87	47,68	89,14	90,68	93,76	62,49
Clasificación bajo nivel Sistema 3	97,93	32,68	81,05	84,32	88,69	47,11
Clasificación nivel de blob Sistema 4	96,54	45,42	90,80	70,18	93,58	55,15
Modelado de frente Sistema 5	96,69	49,83	92,19	71,07	94,39	58,58

Tabla 7-14. Resultados comparativos sobre la secuencia S3 con modelo de fondo de 5 capas.

A continuación se muestran los resultados visuales por medio de diagramas de barras de la Tabla 7-13.

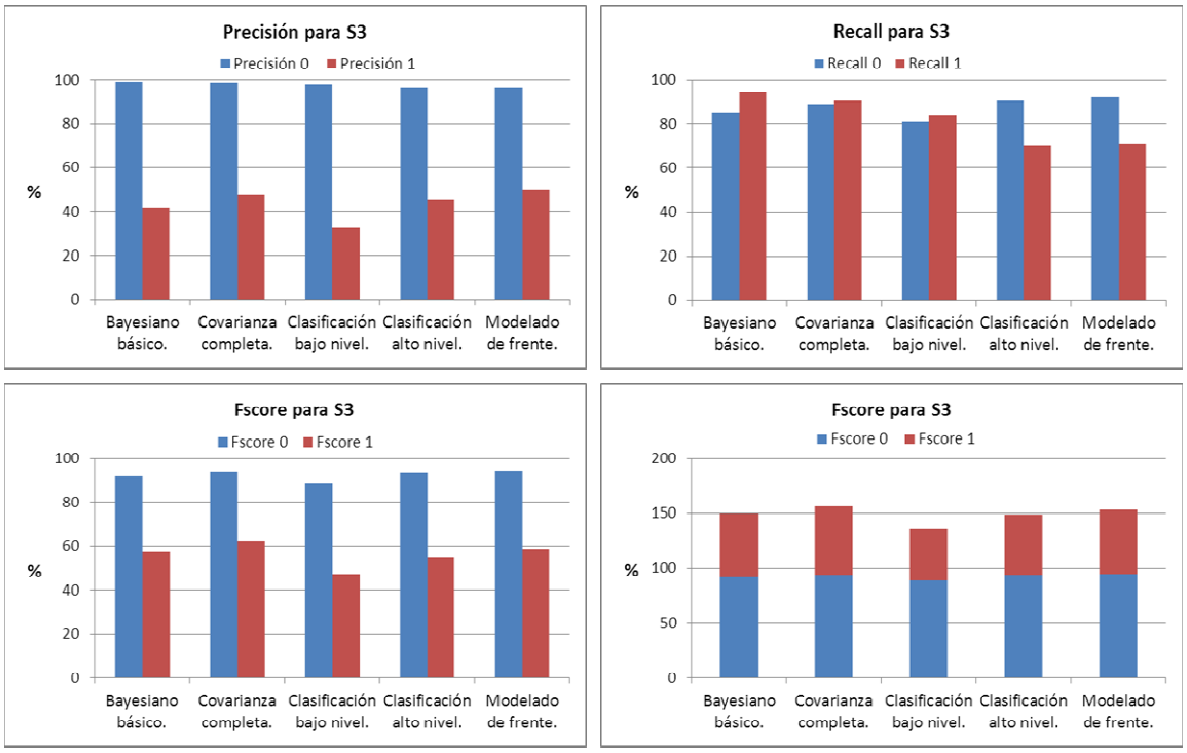


Figura 7-4: Diagramas de barras de los estadísticos de S3.

Resultados cualitativos.

Presentados para el cuadro 150 de la secuencia S3:

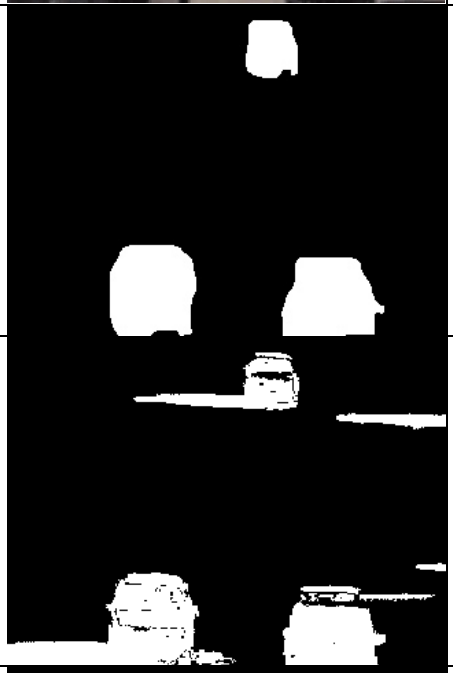
Cuadro			**Fondo**
GT			**Sistema1**
Sistema2			**Sistema3**
Sistema4			**Sistema5**

Tabla 7-15. Resultados cualitativos sobre la secuencia S3.

Discusión.

Los resultados obtenidos para esta secuencia son mejorables y muy parecidos a los de S2. Esto es debido a que en esta secuencia también hay sombras y como se ha explicado anteriormente, los sistemas implementados en este proyecto fin de carrera no son robustos a las sombras.

Como se observa en la Tabla 7-13 y Tabla 7-14 o gráficamente en Figura 7-4, el sistema que mejores resultados ofrece para esta secuencia es el sistema *covarianza completa.* En este sistema, la detección del fondo está por encima del 90% pero la detección de frente apenas supera el 60 %.

Cabe destacar, que para el cuadro mostrado en los resultados cualitativos, todos los sistemas implementados, salvo el *Sistema 3*, se adaptan perfectamente al inicio en caliente del video. Situación que prueba empíricamente, la incapacidad de este sistema, incluso con un **tiempo de margen** alto, de adaptarse a los *fondos estáticos no modelados,* y motiva el desarrollo *del Sistema 4*.

- **S4: Pasillo.**

Características de las secuencias S4: interior, unimodal, con cambios bruscos de iluminación y de complejidad media.

Resultados cuantitativos.

Presentados por medio de tablas y diagramas de barras.

SECUENCIA S4: Pasillo 3 Capas						
Método de Sustracción de fondo.	**P0** **%**	**P1** **%**	**R0** **%**	**R1** **%**	**FS0** **%**	**FS1** **%**
Bayesiano Básico Sistema 1	99,99	83,38	99,09	99,79	99,54	90,85
Covarianza completa Sistema 2	99,99	87,66	99,36	99,69	99,67	93,29
Clasificación bajo nivel Sistema 3	99,99	52,75	95,89	99,86	97,90	69,03
Clasificación nivel de blob Sistema 4	99,98	59,94	96,94	99,61	98,44	74,84
Modelado de frente Sistema 5	99,99	83,23	99,08	99,76	99,53	90,75

Tabla 7-16. Resultados comparativos sobre la secuencia S4 con modelo de fondo de 3 capas.

SECUENCIA S4: Pasillo 5 Capas						
Método de Sustracción de fondo.	**P0** **%**	**P1** **%**	**R0** **%**	**R1** **%**	**FS0** **%**	**FS1** **%**
Bayesiano Básico Sistema 1	99,99	83,41	99,09	99,79	99,54	90,87
Covarianza completa Sistema 2	99,99	87,66	99,36	99,70	99,67	93,29
Clasificación bajo nivel Sistema 3	99,99	52,75	95,89	99,86	97,90	69,03
Clasificación nivel de blob Sistema 4	99,98	59,94	96,94	99,61	98,44	74,84
Modelado de frente Sistema 5	99,99	83,24	99,08	99,76	99,53	90,75

Tabla 7-17. Resultados comparativos sobre la secuencia S4 con modelo de fondo de 5 capas.

A continuación se muestran los resultados visuales por medio de diagramas de barras de la Tabla 7-16.

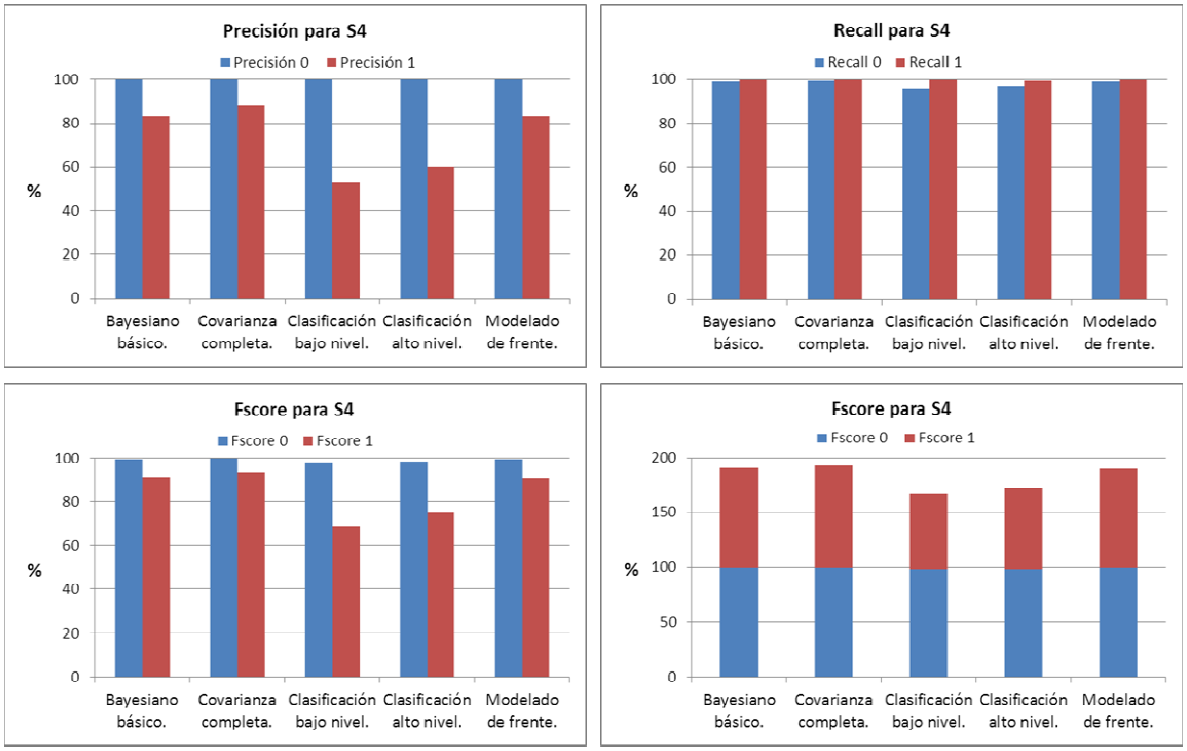


Figura 7-5: Diagramas de barras de los estadísticos de S4.

Resultados cualitativos.

Presentados para el cuadro 1090 de la secuencia S4:

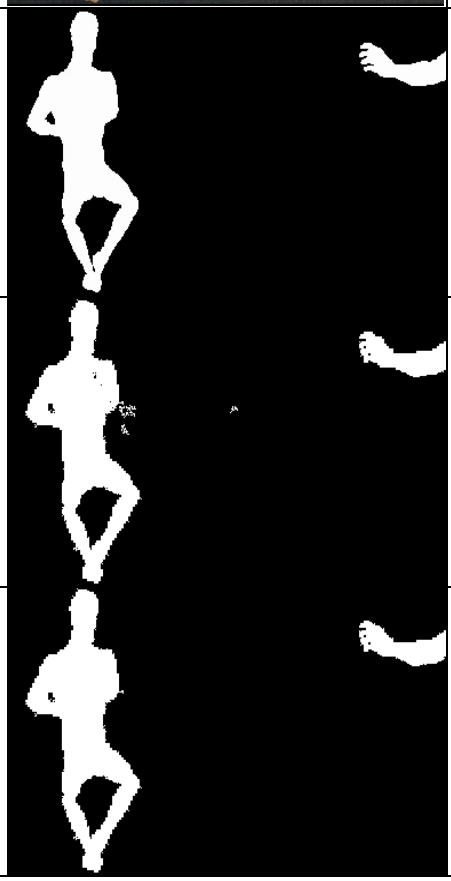
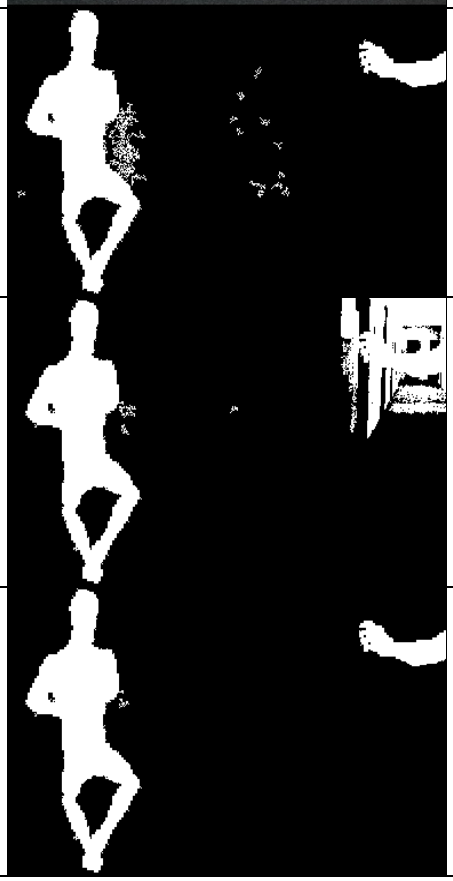
Cuadro			**Fondo**
GT			**Sistema1**
Sistema2			**Sistema3**
Sistema4			**Sistema5**

Tabla 7-18. Resultados cualitativos sobre la secuencia S4.

Discusión.

Los resultados obtenidos para esta secuencia son suficientemente buenos ya que en tres de los sistemas tanto la detección de frente como de fondo está por encima del 90%. En esta secuencia el sistema básico presenta buenos resultados ya que está pensado para este tipo de vídeos: cambios bruscos de iluminación que soporta muy bien y frente que no se repite mucho en la misma zona de la escena evitando que se detecte como fondo.

La clasificación de pixel de alto nivel más modelado de frente (*Sistema 5*) tiene unos resultados muy similares al sistema *Bayesiano Básico* (*Sistema 1*) , ver Tabla 7-16 y Tabla

7-17 . Pero en cambio, no presenta las limitaciones que el modelo básico tiene para vídeos con frentes que permanecen estáticos.

- **S5: Lobby.**

Características de las secuencias S5: interior, unimodal, con cambios bruscos de iluminación, con sombras y de complejidad alta.

Resultados cuantitativos.

Presentados por medio de tablas y diagramas de barras.

SECUENCIA S5: Lobby 3 Capas						
Método de Sustracción de fondo.	**P0** **%**	**P1** **%**	**R0** **%**	**R1** **%**	**FS0** **%**	**FS1** **%**
Bayesiano Básico Sistema 1	99,58	64,83	98,97	82,07	99,27	72,44
Covarianza completa Sistema 2	99,45	70,09	99,25	76,18	99,35	73,01
Clasificación bajo nivel Sistema 3	99,82	13,37	85,94	93,46	92,36	23,39
Clasificación nivel de blob Sistema 4	99,71	66,43	98,97	87,63	99,34	75,57
Modelado de frente Sistema 5	98,84	75,01	99,61	49,88	99,23	59,92

Tabla 7-19. Resultados comparativos sobre la secuencia S5 con modelo de fondo de 3 capas.

SECUENCIA S5: Lobby 5 Capas						
Método de Sustracción de fondo.	**P0** **%**	**P1** **%**	**R0** **%**	**R1** **%**	**FS0** **%**	**FS1** **%**
Bayesiano Básico Sistema 1	99,45	66,49	99,10	76,50	99,28	71,15
Covarianza completa Sistema 2	99,32	71,98	99,36	70,70	99,34	71,33
Clasificación bajo nivel Sistema 3	99,82	13,37	85,94	93,46	92,36	23,39
Clasificación nivel de blob Sistema 4	99,71	66,43	98,97	87,58	99,34	75,55
Modelado de frente Sistema 5	98,84	74,96	99,62	49,45	99,22	59,59

Tabla 7-20. Resultados comparativos sobre la secuencia S5 con modelo de fondo de 5 capas.

A continuación se muestran los resultados visuales por medio de diagramas de barras de la Tabla 7-19.

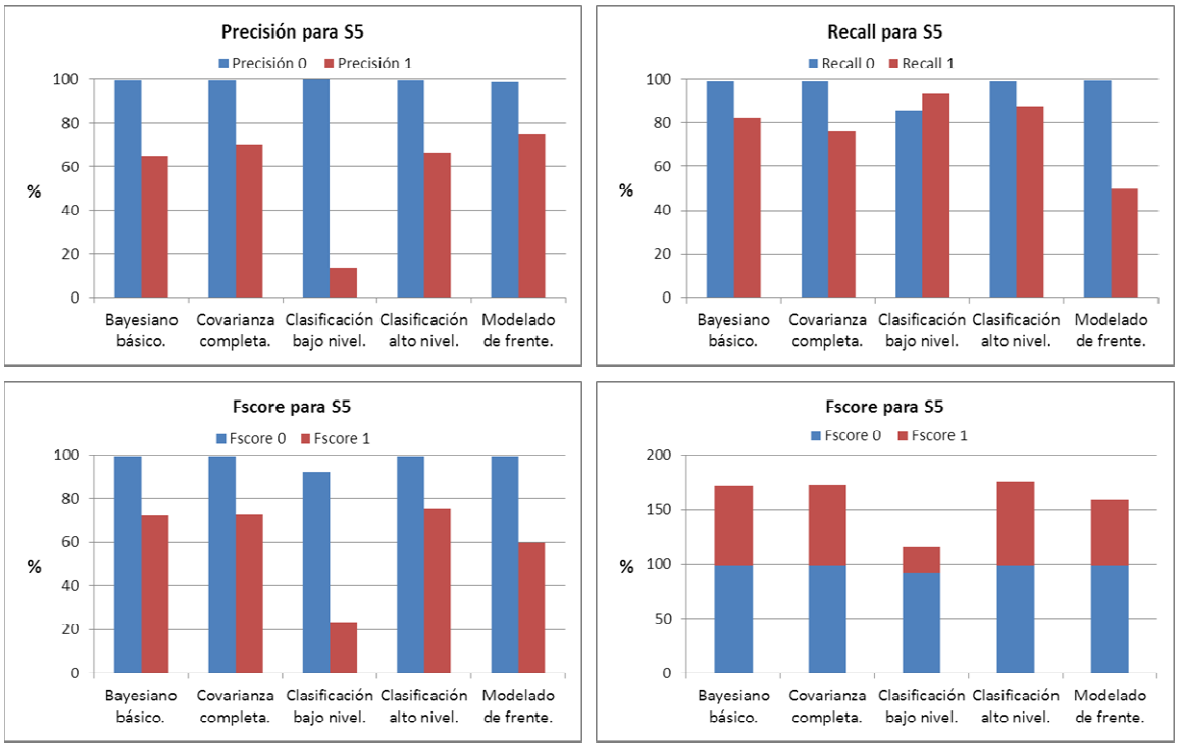


Figura 7-6: Diagramas de barras de los estadísticos de S5.

Resultados cualitativos.

Presentados para el cuadro 1260 de la secuencia S5:

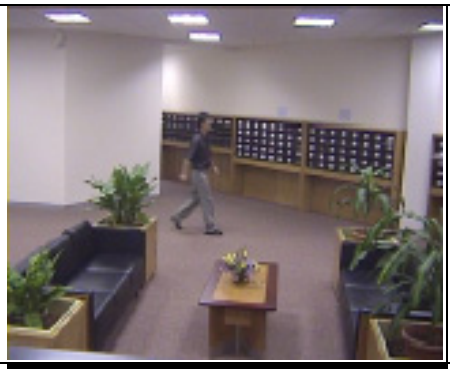
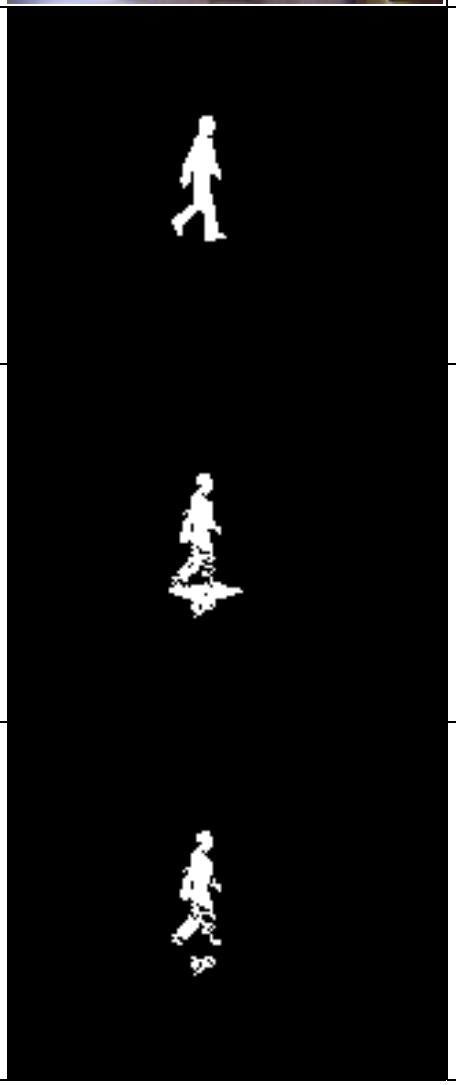
Cuadro			**Fondo**
GT			**Sistema1**
Sistema2			**Sistema3**
Sistema4			**Sistema5**

Tabla 7-21. Resultados cualitativos sobre la secuencia S5.

Discusión.

Los resultados obtenidos en esta secuencia de prueba ofrecen unos buenos resultados para la detección de los píxeles de fondo ya que los cinco sistemas ofrecen resultados por encima del 90% en FS0. Sin embargo los resultados en la detección de frente tienen

margen de mejora ya que en el mejor de los sistemas el FS1 solo se sitúa en un 75% en el mejor de los casos, debido principalmente a la existencia de sombras adheridas a los objetos de frente y de tamaño similar a éstos.

Los mejores resultados para esta secuencia se han conseguido por el *Sistema 4*. (Clasificación de pixel utilizando información de blob). El *Sistema 5* presenta peores resultados que el sistema 4 ya que el modelo de frente incluye píxeles pertenecientes a sombras como píxeles de frente. (Mirar Tabla 7-20).

En esta secuencia se observa una ventaja de los sistemas propuestos que es la buena adaptación que tienen a los cambios bruscos de iluminación (como el global que acontece a la mitad del video) o a los inicios en caliente. (Excepto el *Sistema 3* que no se adapta a los cambios bruscos de esta secuencia). También se observa la desventaja de que ninguno de los sistemas se adapta a las sombras.

- **S6: Ingravidez.**

Características de las secuencias S6: exterior, multimodal alto y de complejidad alta.

Resultados cuantitativos.

Presentados por medio de tablas y diagramas de barras.

SECUENCIA S6: Ingravidez 3 Capas						
Método de Sustracción de fondo.	**P0** **%**	**P1** **%**	**R0** **%**	**R1** **%**	**FS0** **%**	**FS1** **%**
Bayesiano Básico Sistema 1	91,57	32,84	82,12	53,61	86,59	40,73
Covarianza completa Sistema 2	91,29	62,32	95,66	44,04	93,42	51,61
Clasificación bajo nivel Sistema 3	94,98	75,83	96,43	68,73	95,70	72,11
Clasificación nivel de blob Sistema 4	94,56	95,08	99,45	64,90	96,94	77,14
Modelado de frente Sistema 5	95,50	95,45	99,45	71,24	97,43	81,59

Tabla 7-22. Resultados comparativos sobre la secuencia S6 con modelo de fondo de 3 capas.

SECUENCIA S6: Ingravidez 5 Capas						
Método de Sustracción de fondo.	**P0** **%**	**P1** **%**	**R0** **%**	**R1** **%**	**FS0** **%**	**FS1** **%**
Bayesiano Básico Sistema 1	91,48	32,89	82,39	52,93	86,70	40,57
Covarianza completa Sistema 2	91,25	62,61	95,74	43,73	93,44	51,49
Clasificación bajo nivel Sistema 3	94,98	75,83	96,43	68,73	95,70	72,11
Clasificación nivel de blob Sistema 4	94,55	95,11	99,46	64,87	96,94	77,13
Modelado de frente Sistema 5	95,51	95,48	99,45	71,32	97,44	81,65

Tabla 7-23. Resultados comparativos sobre la secuencia S6 con modelo de fondo de 5 capas.

A continuación se muestran los resultados visuales por medio de diagramas de barras de la Tabla 7-22.

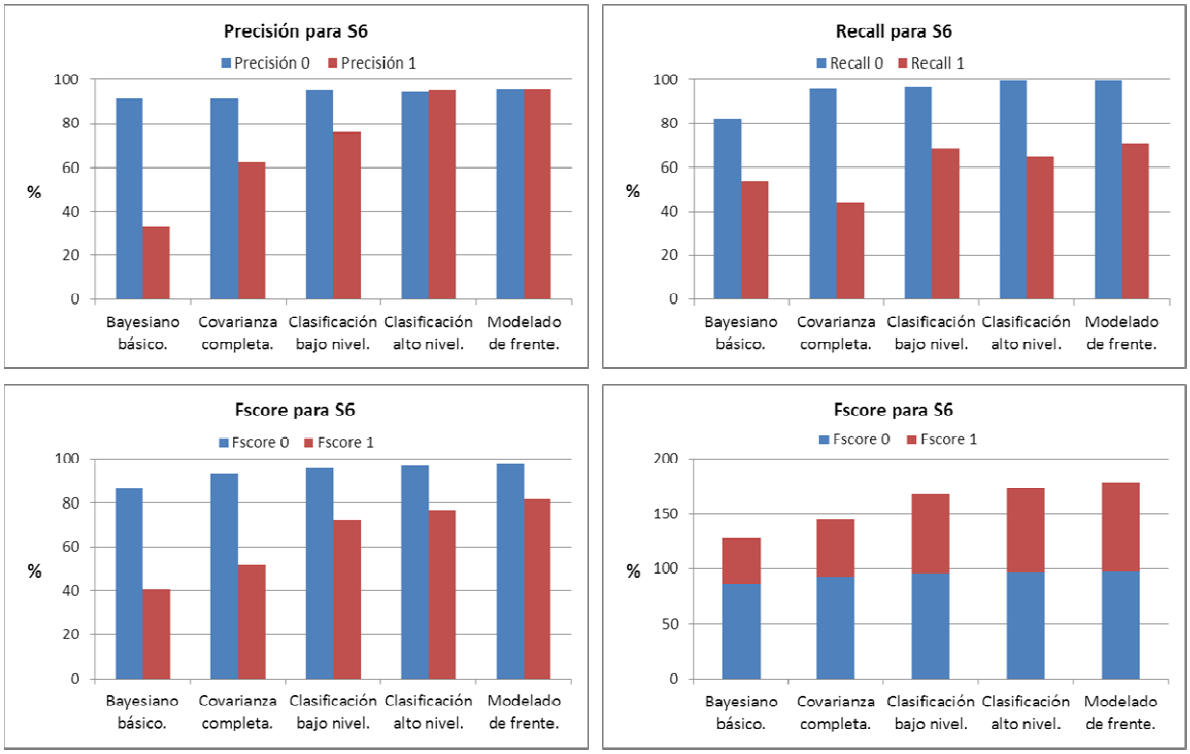


Figura 7-7: Diagramas de barras de los estadísticos de S6.

Resultados cualitativos.

Presentados para el cuadro 241 de la secuencia S6:

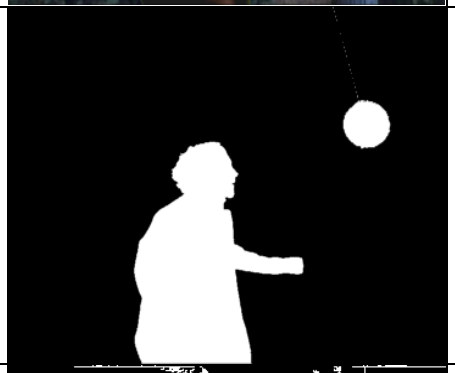
Cuadro			**Fondo**
GT			**Sistema1**
Sistema2			**Sistema3**
Sistema4			**Sistema5**

Tabla 7-24. Resultados cualitativos sobre la secuencia S6.

Discusión.

Los resultados obtenidos en esta secuencia de prueba ofrecen resultados muy buenos tanto para la detección de los píxeles de fondo como para la detección de píxeles de frente. En el *Sistema 5* (modelado de frente) la detección de píxeles de fondo se sitúa por encima

del 97% y la de píxeles de frente por encima del 81%. Esto es debido a que esta secuencia de fondo multimodal y de complejidad muy alta no tiene sombras (característica a la que no se adapta nuestro sistema). Cabe observar que incluso en un video de muy alta modalidad, la aportación de dos capas más no compensa en resultados el alto incremento del coste computacional del sistema.

El sistema final propuesto en este proyecto fin de carrera (*Sistema 5*) ofrece muchas ventajas en la segmentación de este tipo de secuencias. Por un lado es ideal para secuencias multimodales ya que es capaz de modelar diferentes apariencias del pixel en capas aisladas y por otro lado es capaz de detectar píxeles de frente que se camuflarían sin el modelado de frente 6.4. (Esto se puede ver en los resultados cualitativos comparando las máscaras de los *Sistema 4* y *5*).

En la Figura 7-7 se puede ver la mejora constante de los resultados según se van incluyendo nuevas mejoras en los distintos sistemas de segmentación. En los resultados cualitativos se puede apreciar la mejora en las máscaras que se consigue con la inclusión de la clasificación de pixel en el sistema de segmentación.

- **S7: Silla.**

Características de las secuencias S7: exterior, multimodal medio y de complejidad alta.

Resultados cuantitativos.

Presentados por medio de tablas y diagramas de barras.

SECUENCIA S7: Silla 3 Capas						
Método de Sustracción de fondo.	**P0** **%**	**P1** **%**	**R0** **%**	**R1** **%**	**FS0** **%**	**FS1** **%**
Bayesiano Básico Sistema 1	94,68	53,15	93,65	57,79	94,17	55,37
Covarianza completa Sistema 2	93,83	59,92	95,88	49,37	94,84	54,14
Clasificación bajo nivel Sistema 3	97,75	73,38	96,28	82,18	97,01	77,53
Clasificación nivel de blob Sistema 4	97,07	89,89	98,93	76,00	97,99	82,36
Modelado de frente Sistema 5	97,22	89,89	98,92	77,33	98,06	83,14

Tabla 7-25. Resultados comparativos sobre la secuencia S7 con modelo de fondo de 3 capas.

SECUENCIA S7: Silla 5 Capas						
Método de Sustracción de fondo.	**P0** **%**	**P1** **%**	**R0** **%**	**R1** **%**	**FS0** **%**	**FS1** **%**
Bayesiano Básico Sistema 1	94,56	53,38	93,84	56,64	94,19	54,96
Covarianza completa Sistema 2	93,67	60,03	96,03	47,91	94,83	53,29
Clasificación bajo nivel Sistema 3	97,75	73,38	96,29	82,18	97,01	77,53
Clasificación nivel de blob Sistema 4	97,07	90,18	98,97	75,99	98,01	82,48
Modelado de frente Sistema 5	97,22	90,21	98,95	77,31	98,08	83,27

Tabla 7-26. Resultados comparativos sobre la secuencia S7 con modelo de fondo de 5 capas.

A continuación se muestran los resultados visuales por medio de diagramas de barras de la Tabla 7-25.

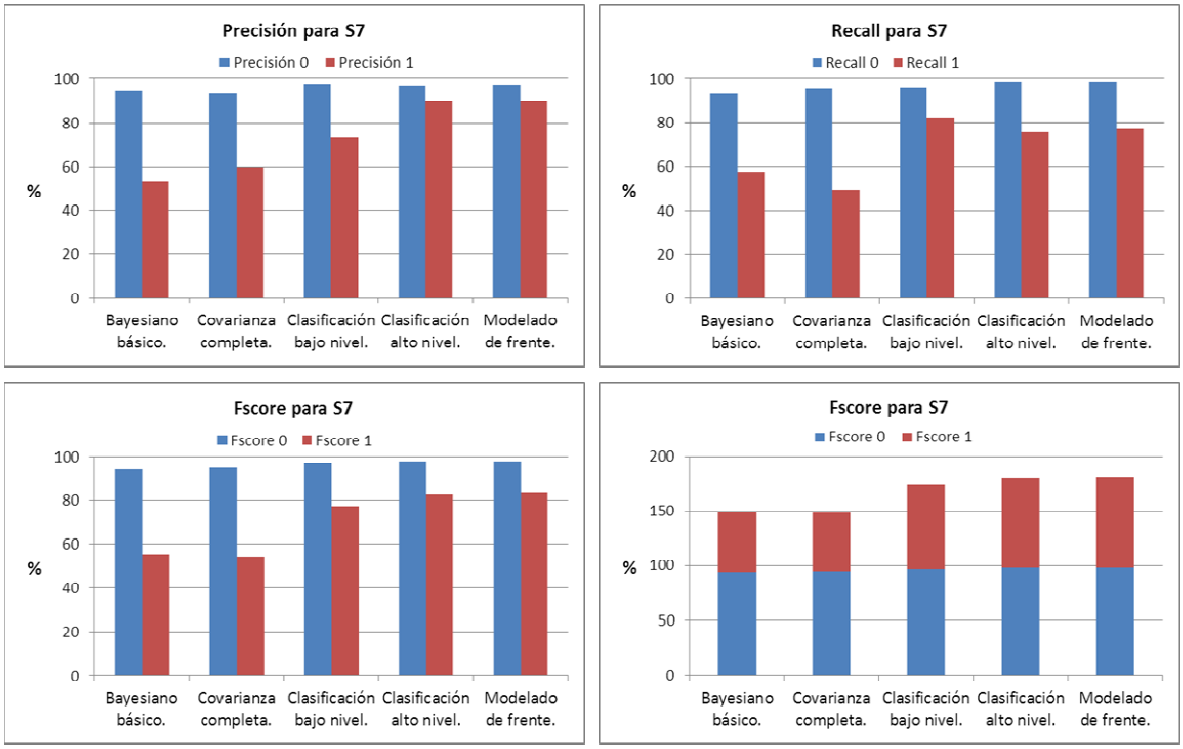


Figura 7-8: Diagramas de barras de los estadísticos de S7.

Resultados cualitativos.

Presentados para el cuadro 420 de la secuencia S7:

Cuadro			**Fondo**
GT			**Sistema1**
Sistema2			**Sistema3**
Sistema4			**Sistema5**

Tabla 7-27. Resultados cualitativos sobre la secuencia S7.

Discusión.

Los resultados obtenidos en esta secuencia de prueba al igual que en la secuencia anterior, muestran resultados muy buenos tanto para la detección de los píxeles de fondo como para la detección de píxeles de frente. En este tipo de secuencias multimodales se

observan las ventajas de modelar el fondo con más capas, aunque la mejora en los estadísticos es pequeña con respecto al aumento del coste computacional requerido.

En el *Sistema 5* (modelado de frente) la detección de píxeles de fondo se sitúa por encima del 98% y la de píxeles de frente por encima del 83%. Esto es debido a que esta secuencia de fondo multimodal y de complejidad muy alta no tiene sombras (característica a la que no se adapta nuestro sistema).

En los resultados cualitativos Tabla 7-27, se observa que la máscara de *ground-truth* perjudica el cálculo de los estadísticos ya que considera a la silla como parte del frente cuando nosotros la consideramos parte del fondo (creemos que conceptualmente debe de ser fondo).

En esta secuencia se muestra un aspecto importante y es que el árbol multimodal que se encuentra en escena no se modela en el fondo en los sistemas básicos (*Sistema 1* y *2*) cuando éstos modelan todas las apariencias del pixel y normalmente en estos sistemas se modelan mejor los fondos multimodales (en contra partida también pierden más píxeles de frente). Esto es debido a que en esta zona de la escena muy multimodal no se gana la confianza suficiente en las capas para ser consideradas capas de fondo fiable puesto que las muestras de los modos que describen el área están dispersas entre sí, generando una varianza muy alta en la descripción del modo (ver sección 5.3.1 y Anexo A).

➢ S8: Hambre.

Características de las secuencias S8: exterior, multimodal bajo y de complejidad baja.

Resultados cuantitativos.

Presentados por medio de tablas y diagramas de barras.

SECUENCIA S8: Hambre 3 Capas						
Método de Sustracción de fondo.	**P0** **%**	**P1** **%**	**R0** **%**	**R1** **%**	**FS0** **%**	**FS1** **%**
Bayesiano Básico Sistema 1	96,49	88,14	99,28	59,62	97,87	71,13
Covarianza completa Sistema 2	96,27	89,77	99,42	56,96	97,82	69,69
Clasificación bajo nivel Sistema 3	99,86	89,01	98,91	98,48	99,39	93,51
Clasificación nivel de blob Sistema 4	99,86	93,32	99,37	98,44	99,61	95,81
Modelado de frente Sistema 5	99,86	93,31	99,37	98,43	99,61	95,80

Tabla 7-28. Resultados comparativos sobre la secuencia S8 con modelo de fondo de 3 capas.

SECUENCIA S8: Hambre 5 Capas						
Método de Sustracción de fondo.	**P0** **%**	**P1** **%**	**R0** **%**	**R1** **%**	**FS0** **%**	**FS1** **%**
Bayesiano Básico Sistema 1	96,42	88,31	99,30	58,77	97,84	70,57
Covarianza completa Sistema 2	96,14	89,84	99,44	55,36	97,76	68,50
Clasificación bajo nivel Sistema 3	99,86	89,01	98,91	98,48	99,39	93,51
Clasificación nivel de blob Sistema 4	99,86	93,32	99,37	98,44	99,61	95,81
Modelado de frente Sistema 5	99,86	93,31	99,37	98,43	99,61	95,80

Tabla 7-29. Resultados comparativos sobre la secuencia S8 con modelo de fondo de 5 capas.

A continuación se muestran los resultados visuales por medio de diagramas de barras de la Tabla 7-28.

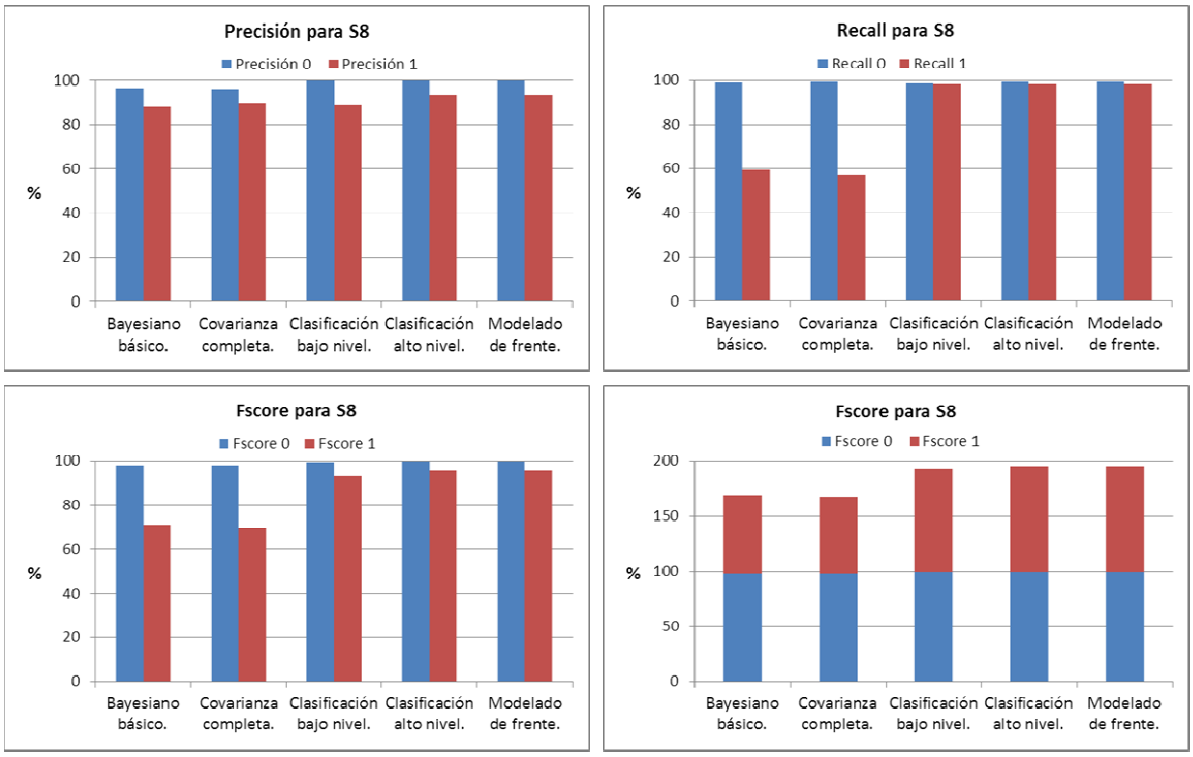


Figura 7-9: Diagramas de barras de los estadísticos de S8.

Resultados cualitativos.

Presentados para el cuadro 748 de la secuencia S8:

Cuadro			**Fondo**
GT			**Sistema1**
Sistema2			**Sistema3**
Sistema4			**Sistema5**

Tabla 7-30. Resultados cualitativos sobre la secuencia S8.

Discusión.

En esta secuencia, los resultados son buenos y se observa la evolución constante en la mejora de los resultados conseguida con la inclusión de cada una de las mejoras. Esto es debido a que es una secuencia sencilla que no presenta demasiados problemas.

En esta secuencia la inclusión del modelado de frente no aporta mejores resultados que la clasificación a nivel de blob ya que el objeto de frente no presenta camuflaje con el fondo.

Una de las desventajas de trabajar a nivel de blob que se aprecia en el cuadro presentado en los resultados cualitativos (Tabla 7-30). Donde se observa que píxeles pertenecientes a un objeto de frente y píxeles de fondo (en otro plano de la imagen, pero que proyectados en el plano 2D se fusionan con los del frente) forman una única componente conexa. En este caso el blob solo puede ser clasificado en una clase (p.e. *frente*) clasificando erróneamente todos los píxeles de una de los dos partes de las que se compone el blob.

- **S9: Jardín.**

Características de las secuencias S9: exterior, multimodal alto y de complejidad alta.

Resultados cuantitativos.

Presentados por medio de tablas y diagramas de barras.

SECUENCIA S9: Jardín 3 Capas						
Método de Sustracción de fondo.	**P0** **%**	**P1** **%**	**R0** **%**	**R1** **%**	**FS0** **%**	**FS1** **%**
Bayesiano Básico Sistema 1	99,82	15,01	88,33	92,66	93,72	25,83
Covarianza completa Sistema 2	99,80	28,36	94,87	91,35	97,27	43,28
Clasificación bajo nivel Sistema 3	99,88	28,57	94,72	94,95	97,23	43,93
Clasificación nivel de blob Sistema 4	99,88	46,43	97,58	94,55	98,71	62,28
Modelado de frente Sistema 5	99,88	47,12	97,64	94,55	98,75	62,89

Tabla 7-31. Resultados comparativos sobre la secuencia S9 con modelo de fondo de 3 capas.

SECUENCIA S9: Jardín 5 Capas						
Método de Sustracción de fondo.	**P0** **%**	**P1** **%**	**R0** **%**	**R1** **%**	**FS0** **%**	**FS1** **%**
Bayesiano Básico Sistema 1	99,80	15,14	88,51	92,20	93,82	26,01
Covarianza completa Sistema 2	99,79	28,55	94,93	91,11	97,30	43,47
Clasificación bajo nivel Sistema 3	99,88	28,57	94,72	94,95	97,23	43,93
Clasificación nivel de blob Sistema 4	99,88	47,32	97,66	94,51	98,76	63,06
Modelado de frente Sistema 5	99,88	47,79	97,70	94,52	98,78	63,48

Tabla 7-32. Resultados comparativos sobre la secuencia S9 con modelo de fondo de 5 capas.

A continuación se muestran los resultados visuales por medio de diagramas de barras de la Tabla 7-31.

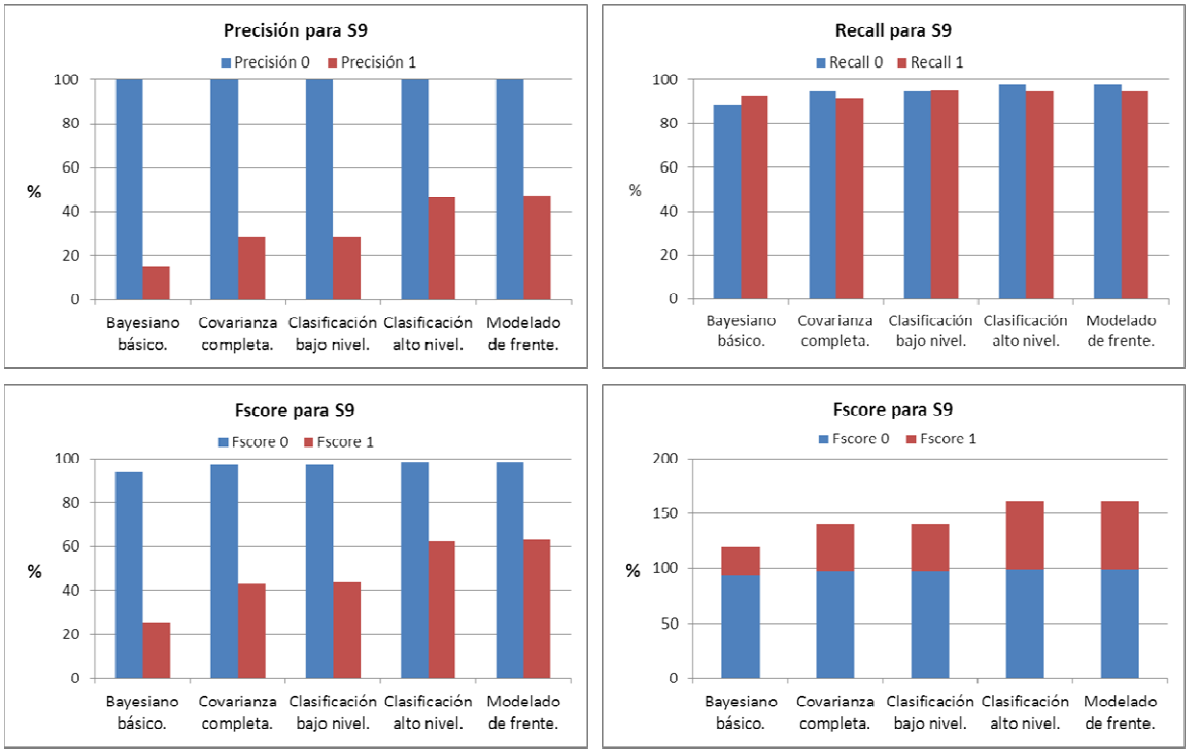


Figura 7-10: Diagramas de barras de los estadísticos de S9.

Resultados cualitativos.

Presentados para el cuadro 555 de la secuencia S9:

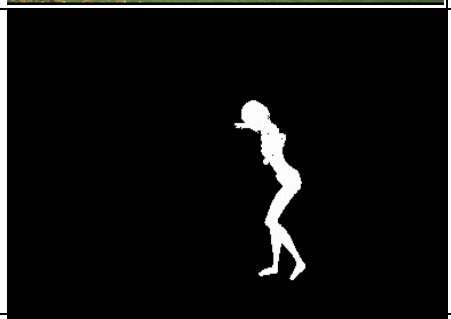
Cuadro			**Fondo**
GT			**Sistema1**
Sistema2			**Sistema3**
Sistema4			**Sistema5**

Tabla 7-33. Resultados cualitativos sobre la secuencia S9.

Discusión.

En esta secuencia se tiene un amplio margen de mejora ya que la detección de frente se sitúa solo en el 63% en el sistema que presenta mejores resultados. Esto se produce porque se detecta como frente una gran cantidad de píxeles pertenecientes a fondos multimodales. Cabe la posibilidad, de que los resultados en este vídeo serían mejores si se hubiese utilizado un umbral mayor (fijado a mano por el experto) a la hora de hacer la discriminación frente fondo en vez de utilizar la umbralización automática de la distancia.

En esta secuencia se ve también una evolución progresiva de los resultados que van mejorando con la inclusión de cada una de las mejoras. Los resultados de la configuración con 5 capas son mejores que la de 3 capas ya que es un fondo muy multimodal pero como en el resto de secuencias la mejora es muy pequeña para el coste computacional requerido.

- **S10: Baile Multimodal.**

Características de las secuencias S10: exterior, multimodal alto y de complejidad alta.

Resultados cuantitativos.

Presentados por medio de tablas y diagramas de barras.

SECUENCIA S10: Baile Multimodal 3 Capas						
Método de Sustracción de fondo.	**P0** **%**	**P1** **%**	**R0** **%**	**R1** **%**	**FS0** **%**	**FS1** **%**
Bayesiano Básico Sistema 1	97,50	27,81	77,03	81,76	86,07	41,51
Covarianza completa Sistema 2	97,22	36,77	85,60	77,35	91,04	49,85
Clasificación bajo nivel Sistema 3	99,15	54,43	91,59	92,77	95,22	68,61
Clasificación nivel de blob Sistema 4	99,01	77,56	97,15	91,03	98,07	83,76
Modelado de frente Sistema 5	99,36	80,84	97,58	94,21	98,46	87,01

Tabla 7-34. Resultados comparativos sobre la secuencia S10 con modelo de fondo de 3 capas.

SECUENCIA S10: Baile Multimodal 5 Capas						
Método de Sustracción de fondo.	**P0** **%**	**P1** **%**	**R0** **%**	**R1** **%**	**FS0** **%**	**FS1** **%**
Bayesiano Básico Sistema 1	97,43	27,76	77,13	81,21	86,10	41,38
Covarianza completa Sistema 2	97,09	36,66	85,74	76,23	91,06	49,51
Clasificación bajo nivel Sistema 3	99,15	54,43	91,59	92,77	95,22	68,61
Clasificación nivel de blob Sistema 4	99,01	77,62	97,16	91,00	98,07	83,78
Modelado de frente Sistema 5	99,35	80,02	97,46	94,11	98,39	86,49

Tabla 7-35. Resultados comparativos sobre la secuencia S10 con modelo de fondo de 5 capas.

A continuación se muestran los resultados visuales por medio de diagramas de barras de la Tabla 7-34.

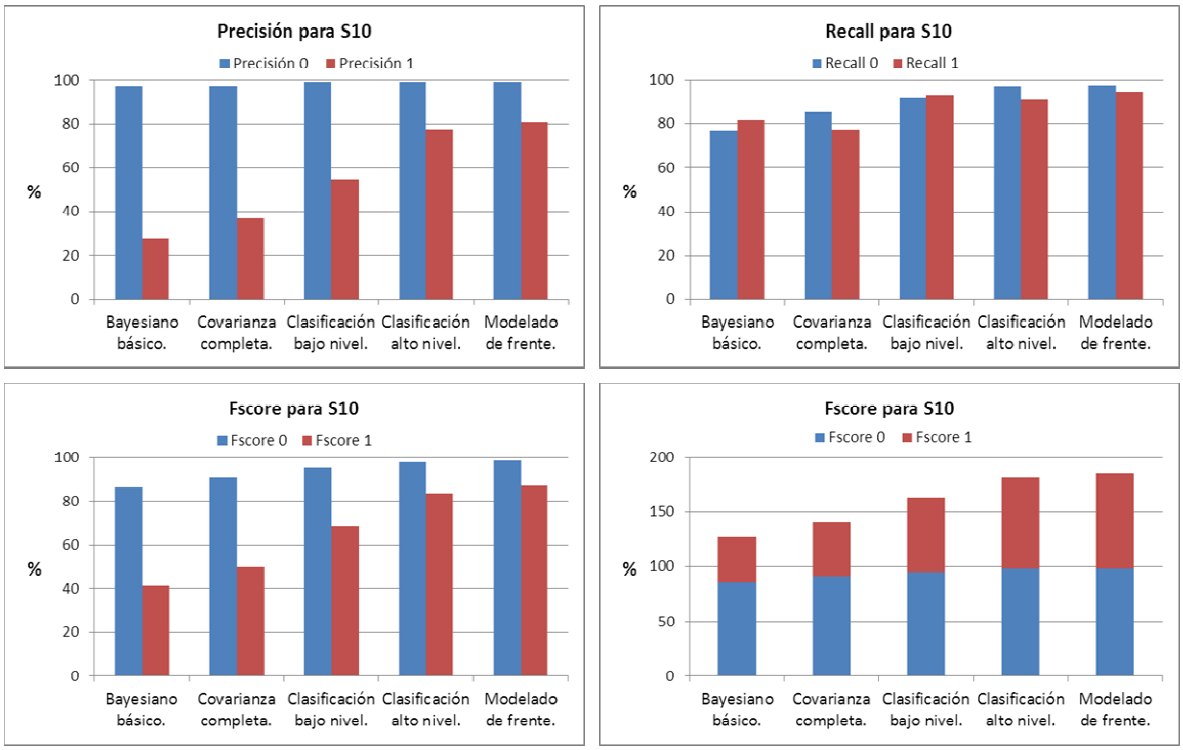


Figura 7-11: Diagramas de barras de los estadísticos de S10.

Resultados cualitativos.

Presentados para el cuadro 532 de la secuencia S10:

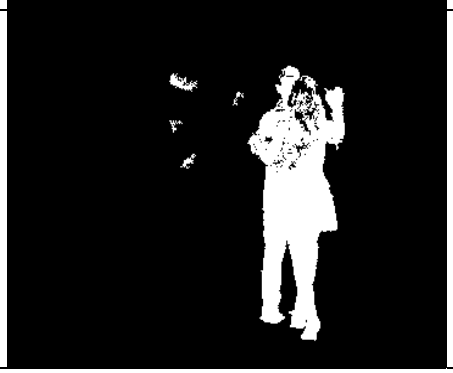
Cuadro			**Fondo**
GT			**Sistema1**
Sistema2			**Sistema3**
Sistema4			**Sistema5**

Tabla 7-36. Resultados cualitativos sobre la secuencia S10.

Discusión.

Los resultados conseguidos en esta secuencia son muy buenos ya que se trata de una secuencia de fondos muy multimodales sin sombras en escena. Características ideales para el tipo de segmentador que se ha diseñado.

En los resultados se nota la inclusión de cada una de las mejoras. Los resultados del *Sistema 2* con respecto al *Sistema 1* demuestran la mejora en la discriminación del uso de la matriz de covarianza. Por otro lado los resultados del *Sistema 3* y del *Sistema 4* demuestran que la clasificación de pixel logra que se pierdan menos píxeles pertenecientes al frente y que se modelen mejor los fondos multimodales. Por último los resultados del sistema 5 demuestran que el modelado de frente logra recuperar algunos píxeles que se perdían por camuflaje con el fondo.

La mayoría de las hipótesis formuladas a lo largo de este documento se observan con claridad en Tabla 7-36.

➢ S11: Curtain.

Características de las secuencias S11: interior, multimodal medio, con cambios bruscos de iluminación, con sombras y de complejidad alta.

Resultados cuantitativos.

Presentados por medio de tablas y diagramas de barras.

SECUENCIA S11: Curtain 3 Capas						
Método de Sustracción de fondo.	**P0** **%**	**P1** **%**	**R0** **%**	**R1** **%**	**FS0** **%**	**FS1** **%**
Bayesiano Básico Sistema 1	98,64	83,35	98,31	86,19	98,47	84,75
Covarianza completa Sistema 2	97,48	89,06	99,11	73,91	98,29	80,78
Clasificación bajo nivel Sistema 3	99,22	31,50	79,99	93,57	88,57	47,13
Clasificación nivel de blob Sistema 4	98,93	60,68	94,29	89,67	96,56	72,38
Modelado de frente Sistema 5	99,08	58,86	93,74	91,15	96,33	71,53

Tabla 7-37. Resultados comparativos sobre la secuencia S11 con modelo de fondo de 3 capas.

SECUENCIA S11: Curtain 5 Capas						
Método de Sustracción de fondo.	**P0** **%**	**P1** **%**	**R0** **%**	**R1** **%**	**FS0** **%**	**FS1** **%**
Bayesiano Básico Sistema 1	97,74	85,86	98,76	76,77	98,25	81,06
Covarianza completa Sistema 2	95,97	87,08	99,16	57,67	97,54	69,39
Clasificación bajo nivel Sistema 3	99,22	31,50	79,99	93,57	88,57	47,13
Clasificación nivel de blob Sistema 4	98,93	60,83	94,33	89,62	96,57	72,47
Modelado de frente Sistema 5	99,08	59,04	93,78	91,13	96,36	71,66

Tabla 7-38. Resultados comparativos sobre la secuencia S11 con modelo de fondo de 5 capas.

A continuación se muestran los resultados visuales por medio de diagramas de barras de la Tabla 7-37.

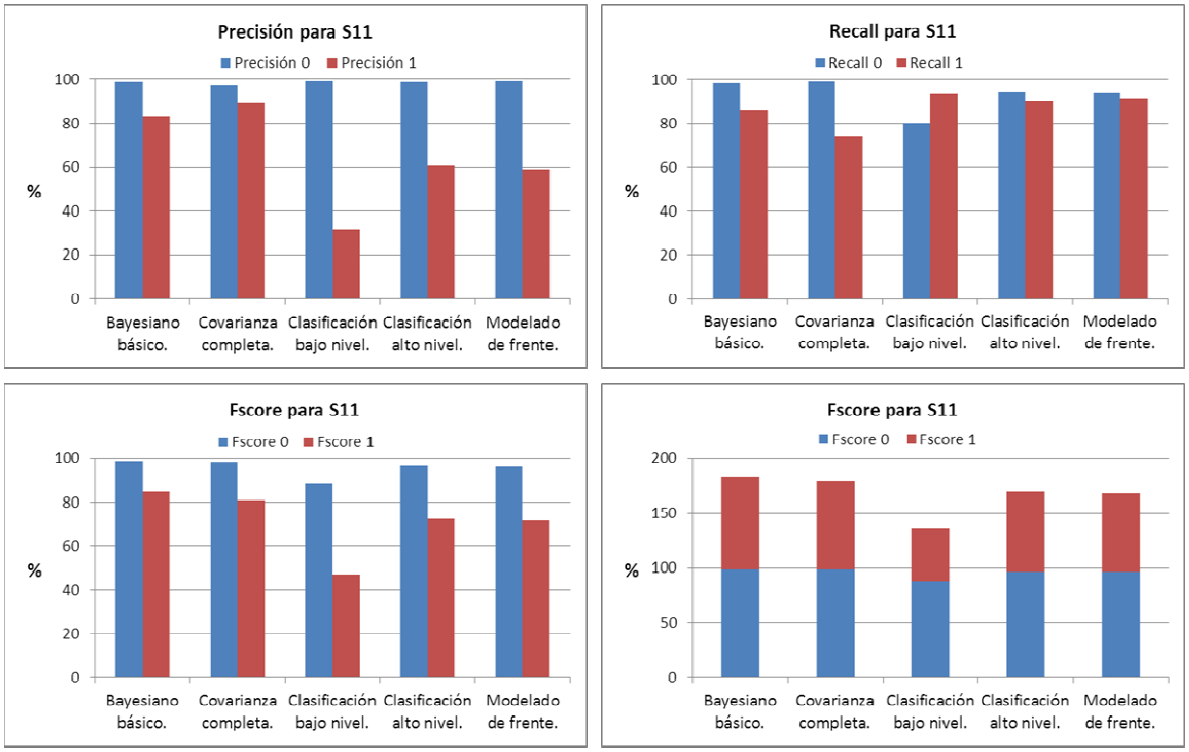


Figura 7-12: Diagramas de barras de los estadísticos de S11.

Resultados cualitativos.

Presentados para el cuadro 1847 de la secuencia S11:

Cuadro			**Fondo**
GT			**Sistema1**
Sistema2			**Sistema3**
Sistema4			**Sistema5**

Tabla 7-39. Resultados cualitativos sobre la secuencia S11.

Discusión.

Los resultados obtenidos en esta secuencia no son buenos para las mejoras introducidas ya que los mejores resultados se consiguen con el *Sistema Básico* (*Sistema 1*). Esto es debido a que la secuencia tiene las características donde mejor se adapta el *Sistema*

1. Fondos multimodales con cambios bruscos de iluminación y donde los objetos de frente no se repiten mucho en la misma posición de la escena.

También han influido negativamente en los resultados los reflejos y las sombras producidas en la secuencia que formaban componentes conexas con los objetos de frente.

Puesto que esta secuencia es para una de las que peor funciona el sistema (atendiendo a su complejidad) se ha seleccionado como candidata para realizar un estudio exhaustivo de la influencia de la inicialización de los parámetros de análisis (Influencia de la inicialización de σ_0 .7.3.4).

- **S12: Water Surface.**

Características de las secuencias S12: exterior, multimodal alto y de complejidad alta.

Resultados cuantitativos.

Presentados por medio de tablas y diagramas de barras.

SECUENCIA S12: Water Surface 3 Capas						
Método de Sustracción de fondo.	**P0** **%**	**P1** **%**	**R0** **%**	**R1** **%**	**FS0** **%**	**FS1** **%**
Bayesiano Básico Sistema 1	98,93	68,58	96,49	87,98	97,70	77,08
Covarianza completa Sistema 2	98,61	94,53	99,58	83,85	99,09	88,87
Clasificación bajo nivel Sistema 3	99,56	94,08	99,48	94,89	99,52	94,49
Clasificación nivel de blob Sistema 4	99,54	94,22	99,49	94,76	99,52	94,49
Modelado de frente Sistema 5	99,54	94,22	99,49	94,76	99,52	94,49

Tabla 7-40. Resultados comparativos sobre la secuencia S12 con modelo de fondo de 3 capas.

SECUENCIA S12: Water Surface 5 Capas						
Método de Sustracción de fondo.	**P0** **%**	**P1** **%**	**R0** **%**	**R1** **%**	**FS0** **%**	**FS1** **%**
Bayesiano Básico Sistema 1	98,94	69,69	96,67	88,11	97,79	77,83
Covarianza completa Sistema 2	98,63	94,51	99,57	84,13	99,10	89,02
Clasificación bajo nivel Sistema 3	99,56	94,08	99,48	94,89	99,52	94,49
Clasificación nivel de blob Sistema 4	99,54	94,22	99,49	94,76	99,52	94,49
Modelado de frente Sistema 5	99,54	94,22	99,49	94,76	99,52	94,49

Tabla 7-41. Resultados comparativos sobre la secuencia S12 con modelo de fondo de 5 capas.

A continuación se muestran los resultados visuales por medio de diagramas de barras de la Tabla 7-40.

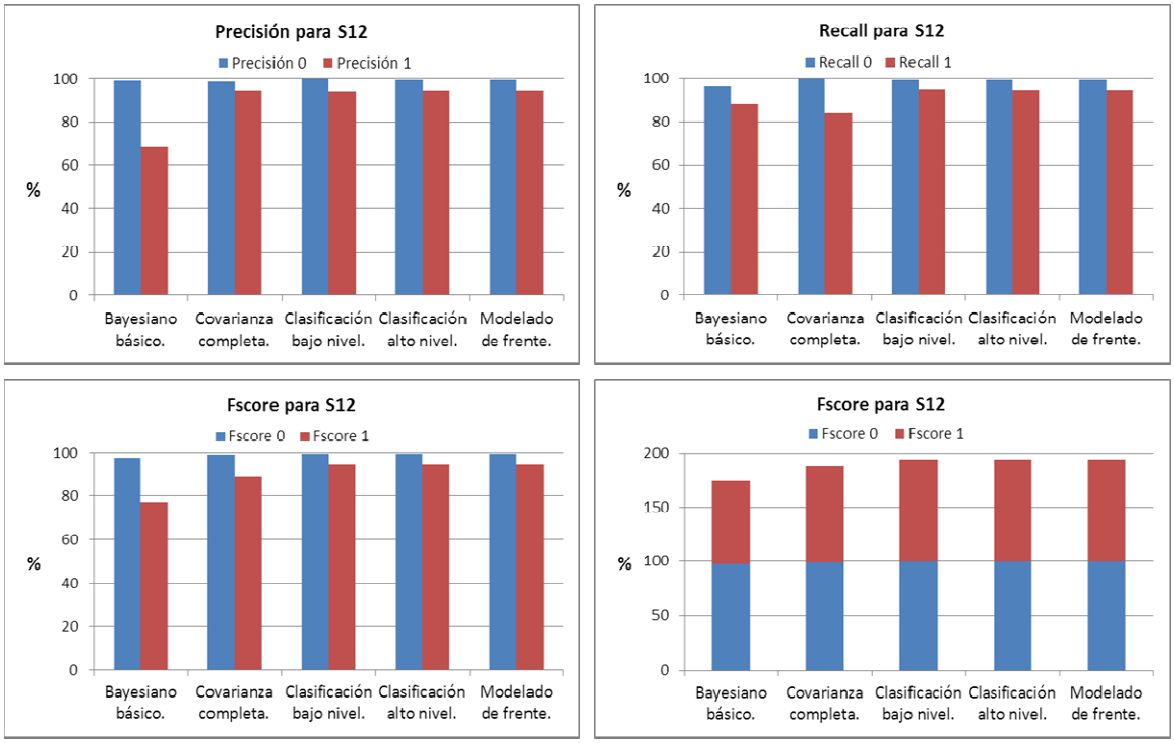


Figura 7-13: Diagramas de barras de los estadísticos de S12.

Resultados cualitativos.

Presentados para el cuadro 620 de la secuencia S12:

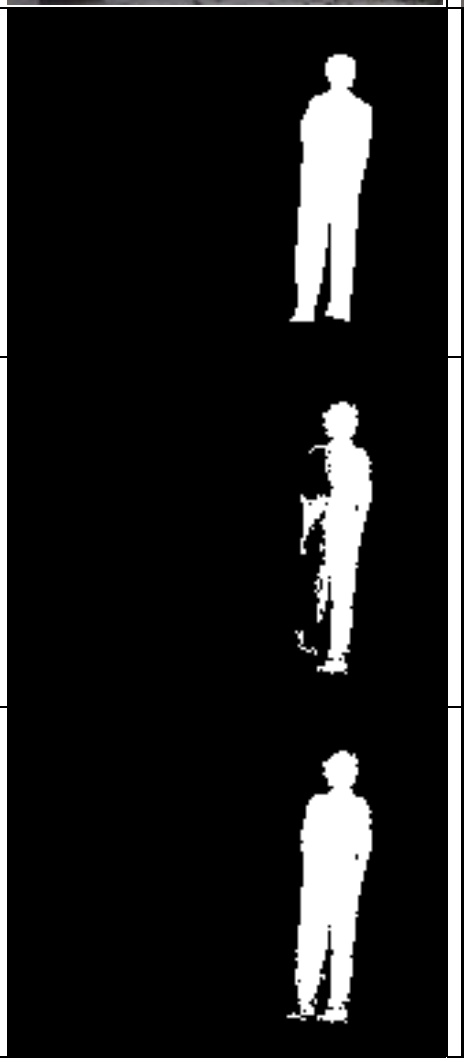
Cuadro		**Fondo**
GT		**Sistema1**
Sistema2		**Sistema3**
Sistema4		**Sistema5**

Tabla 7-42. Resultados cualitativos sobre la secuencia S12.

Discusión.

Los resultados de esta secuencia han sido muy buenos lo que demuestra que el algoritmo se adapta a distintos fondos multimodales. Los resultados de detección de fondo están por encima del 94 % en los tres sistemas que utilizan clasificación de pixel y por encima del 94 % en detección de frente. Puede observarse (en los *Sistemas 1* y *2*) la

incorrecta clasificación de los píxeles de *frente* por estaticidad del objeto de frente, situación que se solventa en el resto de los sistemas.

➢ S13: Fountain.

Características de las secuencias S13: exterior, multimodal alto, con sombras y de complejidad alta.

Resultados cuantitativos.

Presentados por medio de tablas y diagramas de barras.

SECUENCIA S13: Fountain 3 Capas						
Método de Sustracción de fondo.	**P0** **%**	**P1** **%**	**R0** **%**	**R1** **%**	**FS0** **%**	**FS1** **%**
Bayesiano Básico Sistema 1	99,60	38,34	94,91	89,19	97,19	53,63
Covarianza completa Sistema 2	99,40	81,82	99,34	83,23	99,37	82,52
Clasificación bajo nivel Sistema 3	99,41	40,91	95,69	83,94	97,52	55,01
Clasificación nivel de blob Sistema 4	99,41	72,73	98,89	83,39	99,15	77,70
Modelado de frente Sistema 5	99,44	80,64	99,28	84,15	99,36	82,36

Tabla 7-43. Resultados comparativos sobre la secuencia S13 con modelo de fondo de 3 capas.

SECUENCIA S13: Fountain 5 Capas						
Método de Sustracción de fondo.	**P0** **%**	**P1** **%**	**R0** **%**	**R1** **%**	**FS0** **%**	**FS1** **%**
Bayesiano Básico Sistema 1	99,60	39,92	95,24	89,10	97,37	55,14
Covarianza completa Sistema 2	99,40	81,80	99,34	83,20	99,37	82,50
Clasificación bajo nivel Sistema 3	99,41	40,91	95,69	83,94	97,52	55,01
Clasificación nivel de blob Sistema 4	99,41	72,73	98,89	83,39	99,15	77,70
Modelado de frente Sistema 5	99,44	80,64	99,28	84,14	99,36	82,36

Tabla 7-44. Resultados comparativos sobre la secuencia S13 con modelo de fondo de 5 capas.

A continuación se muestran los resultados visuales por medio de diagramas de barras de la Tabla 7-43.

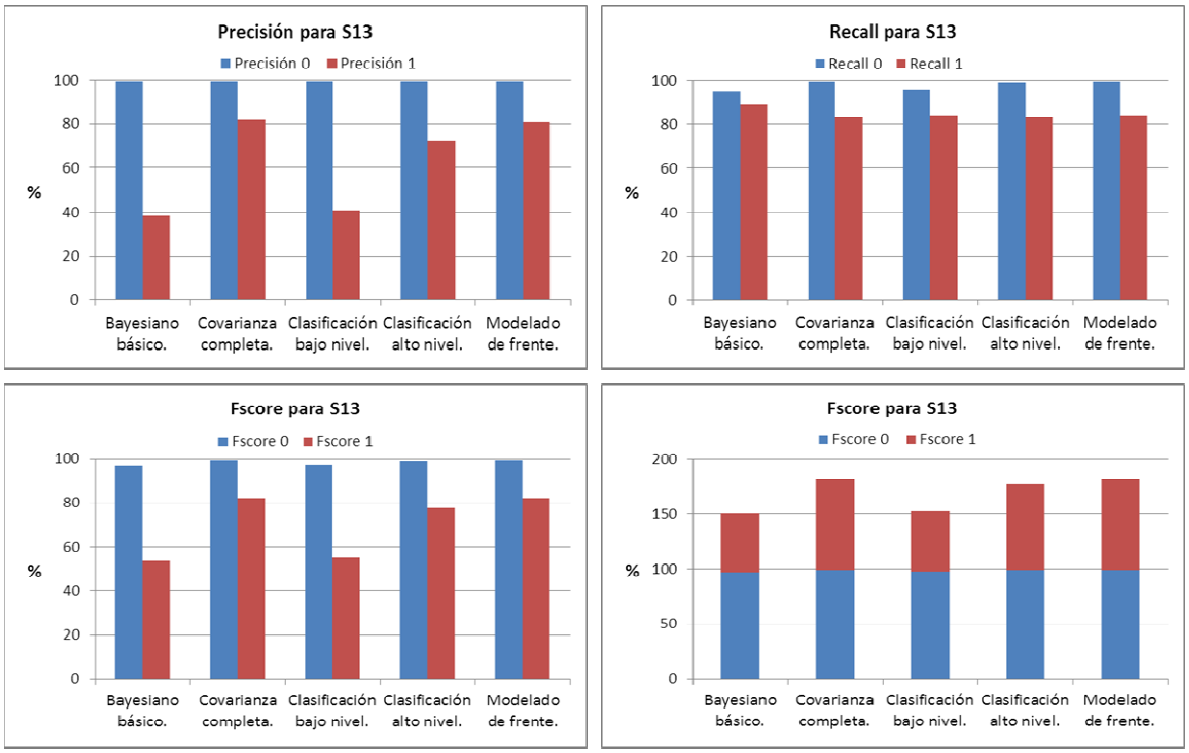


Figura 7-14: Diagramas de barras de los estadísticos de S13.

Resultados cualitativos.

Presentados para el cuadro 509 de la secuencia S13:

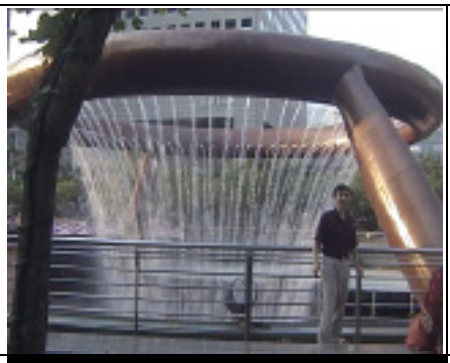
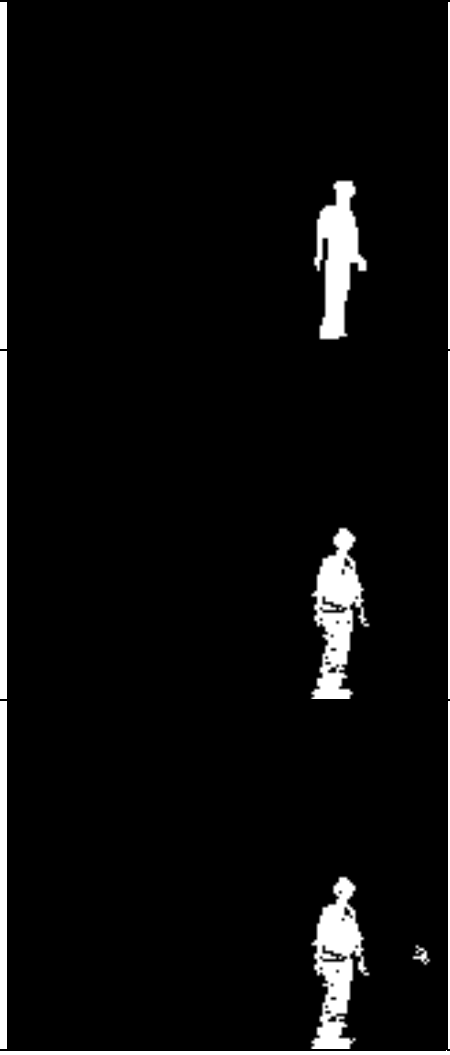
Cuadro			**Fondo**
GT			**Sistema1**
Sistema2			**Sistema3**
Sistema4			**Sistema5**

Tabla 7-45. Resultados cualitativos sobre la secuencia S13.

Discusión.

Esta secuencia tiene una alta complejidad y tiene características (como la sombras) a las que nuestro algoritmo no se adapta bien. Sin embargo los resultados obtenidos con el *Sistema 2* y S*istema 5* han sido aceptable ya que logran una detección de fondo por e cima del 99% y de frente por encima del 82%.

En este video se observa claramente las ventajas de trabajar con covarianza completa (evolución desde el *Sistema 1* al *Sistema 2*) y a nivel de blob (evolución desde el *Sistema 3* al *Sistema 4*).

➢ S14: Escalator.

Características de las secuencias S14: interior, multimodal alto, con sombras, con cambios bruscos de iluminación y de complejidad alta.

Resultados cuantitativos.

Presentados por medio de tablas y diagramas de barras.

SECUENCIA S14: Escalator 3 Capas						
Método de Sustracción de fondo.	**P0** **%**	**P1** **%**	**R0** **%**	**R1** **%**	**FS0** **%**	**FS1** **%**
Bayesiano Básico Sistema 1	99,44	17,72	80,98	89,97	89,26	29,61
Covarianza completa Sistema 2	99,15	24,07	88,02	83,40	93,25	37,36
Clasificación bajo nivel Sistema 3	98,87	25,55	89,72	77,44	94,07	38,42
Clasificación nivel de blob Sistema 4	98,82	48,72	96,42	74,76	97,60	59,00
Modelado de frente Sistema 5	98,80	51,97	96,88	74,11	97,83	61,10

Tabla 7-46. Resultados comparativos sobre la secuencia S14 con modelo de fondo de 3 capas.

SECUENCIA S14: Escalator 5 Capas						
Método de Sustracción de fondo.	**P0** **%**	**P1** **%**	**R0** **%**	**R1** **%**	**FS0** **%**	**FS1** **%**
Bayesiano Básico Sistema 1	99,43	18,21	81,65	89,70	89,66	30,27
Covarianza completa Sistema 2	99,13	24,36	88,27	82,98	93,38	37,67
Clasificación bajo nivel Sistema 3	98,87	25,55	89,72	77,44	94,07	38,42
Clasificación nivel de blob Sistema 4	98,81	50,04	96,62	74,44	97,70	59,85
Modelado de frente Sistema 5	98,79	53,41	97,07	73,84	97,92	61,98

Tabla 7-47. Resultados comparativos sobre la secuencia S14 con modelo de fondo de 5 capas.

A continuación se muestran los resultados visuales por medio de diagramas de barras de la Tabla 7-46.

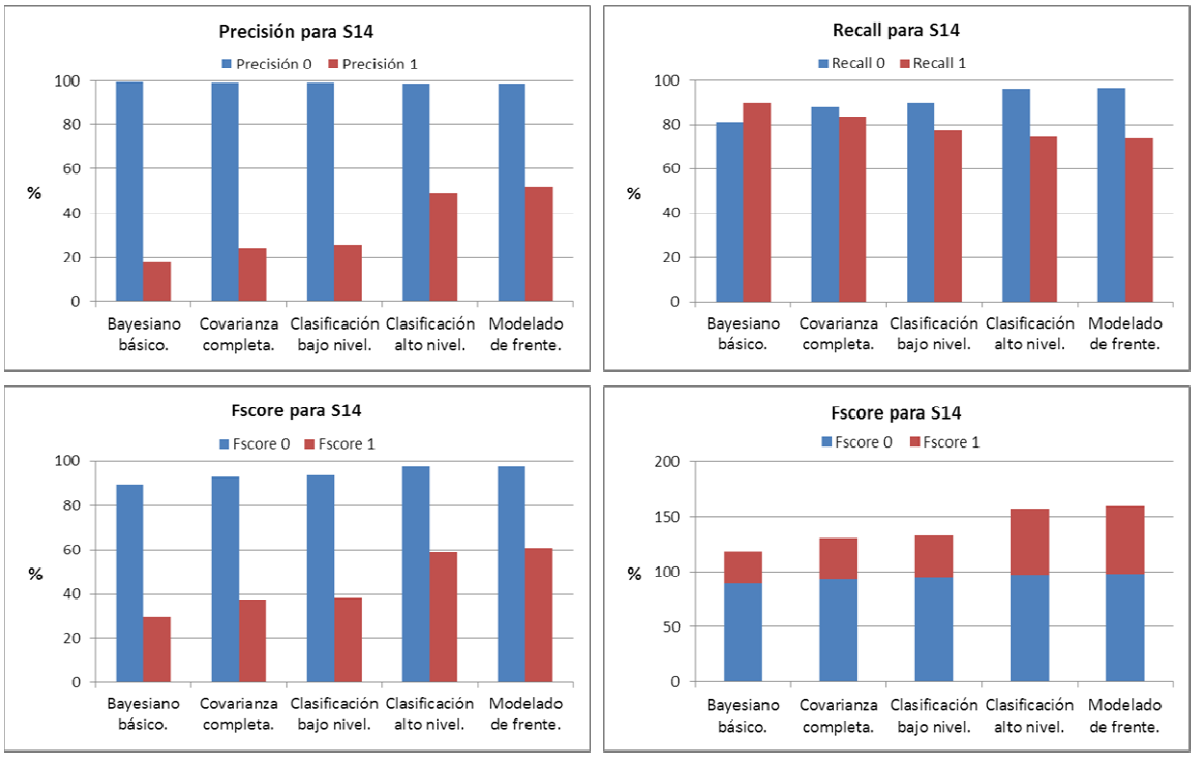


Figura 7-15: Diagramas de barras de los estadísticos de S14.

Resultados cualitativos.

Presentados para el cuadro 3389 de la secuencia S14:

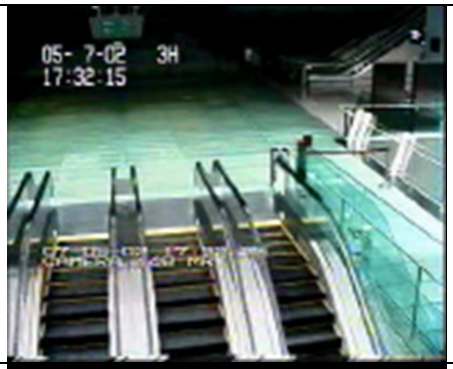
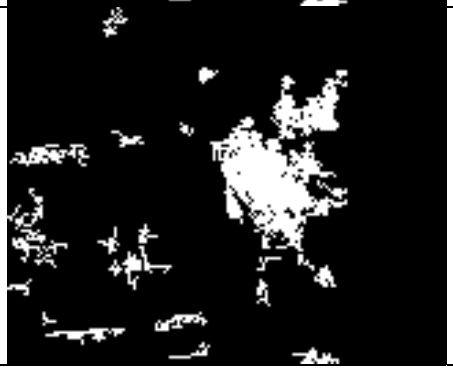
Cuadro			**Fondo**
GT			**Sistema1**
Sistema2			**Sistema3**
Sistema4			**Sistema5**

Tabla 7-48. Resultados cualitativos sobre la secuencia S14.

Discusión.

Esta secuencia tiene numerosos factores de complejidad ya que tiene numerosos cambios de iluminación, un fondo multimodal muy complicado con la presencia de las escaleras mecánicas y un objeto de frente con sombras.

Los resultados presentan margen de mejora. Pero en ellos se ve la evolución que cada una de las mejoras ha producido. El sistema que mejores resultados ofrece es el *Sistema 5* que ofrece una detección de fondo por encima del 97% y de frente por encima del 60%.

Merecen destacarse los resultados obtenidos por el *Sistema 3* en la zona de las escaleras mecánicas; al realizar correctamente la clasificación de sus píxeles en fondo dinámico no modelado, este sistema es más restrictivo que los *Sistemas 1* y *2* en la introducción de nuevas apariencias en el modelo. Debido a esto, la *fdp* que modela el valor de la distancia en las capas de fondo tiene una desviación típica más cercana a la media y por ello confianzas más altas que superan el umbral.

7.3.4 **Influencia de la inicialización de σ_0.**

En esta sección se muestran diferentes curvas ROC resultado de aplicar distintos valores de inicialización al parámetro σ_0 (ver sección 7.2) para los sistemas desarrollados en este proyecto fin de carrera.

Con estas curvas ROC se intenta evaluar la influencia de esta inicialización en los resultados conseguidos. Se muestran los resultados sobre la secuencia S11 (*Curtain*), por ser esta una de las que resultan en peores resultados.

Una curva ROC (*Receiver Operating Characteristic*), es una representación gráfica de la sensibilidad frente a (1 – especificidad). Se define sensibilidad como razón de verdaderos positivos (**R1** definido en ecuación *eq 7.9*) y especificidad como razón de verdaderos negativos (**R0** definido en ecuación *eq 7.8*).

Se define un nuevo parámetro β que es resultado de la indexación de la siguiente rejilla:

$$\beta \in [3.5, 3.25, 3, 2.75, 2.5, 2.25, 2, 1.75, 1.5, 1.25, 1, 0.5, 0.25, 0.1] \qquad \textit{eq 7.13}$$

En función de este parámetro se define la inicialización de σ_0 para los diferentes sistemas implementados en este proyecto.

En los sistemas *Bayesiano Básico (Sistema 1)*, *Bayesiano Básico* con matriz de covarianza completa *(Sistema 2)* y en clasificación a nivel de pixel *(Sistema 3)* la inicialización utilizada es la siguiente:

$$\sigma_0 = 2\beta \qquad \textit{eq 7.14}$$

Para la clasificación de pixel a nivel de blob *(Sistema 4)*, y para el modelado de frente *(Sistema 5)*, la inicialización para los píxeles clasificados como *fondo estático no modelado* es:

$$\sigma_0 = 3\beta \qquad \textit{eq 7.15}$$

Y para los clasificados como *fondo dinámico no modelado*:

$$\sigma_0 = \beta \qquad \textit{eq 7.16}$$

La siguiente figura ilustra los resultados obtenidos mediante curvas ROC para cada sistema y para cada valor de β analizado.

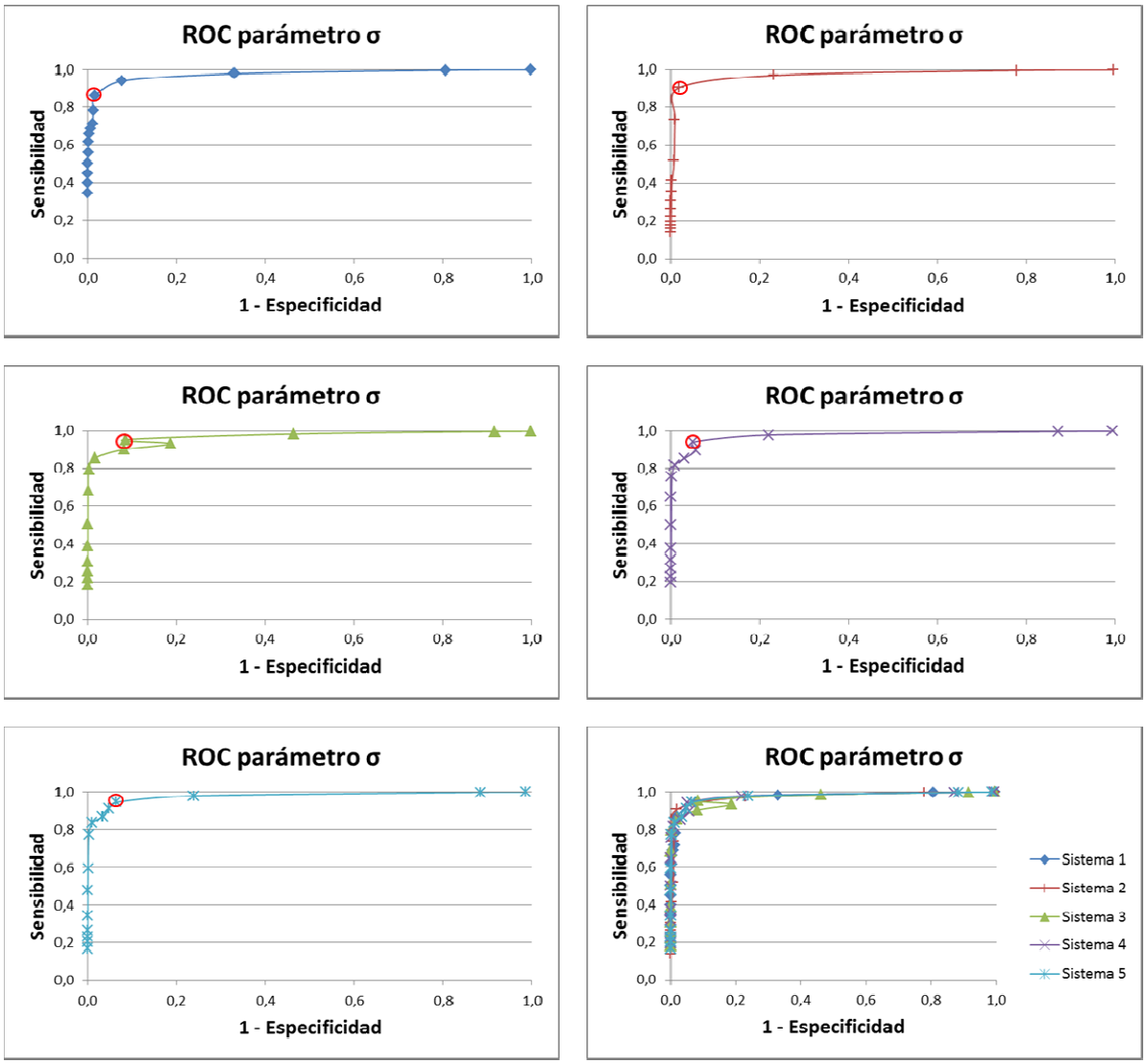


Figura 7-16: Curvas ROC para evaluar la inicialización de σ_0.

Se marca con un círculo rojo la inicialización óptima para cada uno de los sistemas. El parámetro β óptimo de inicialización para cada sistema es: $\beta = 1$ para *Sistema 1*, y $\beta = 0,75$ para *Sistema 2, 3, 4 y 5*. En los resultados mostrados en 7.3.3, la inicialización de todos los sistemas se ha hecho para $\beta = 1$. Esta es una de las razones por la cual en esta secuencia S11, los mejores resultados se consiguen con el *Sistema 1*.

7.3.5 Comparativa con el SoA.

En esta sección se muestran tablas donde se comparan los resultados de este trabajo fin de carrera con los resultados obtenidos por otros autores. Sólo se compararán los

resultados obtenidos para algunos videos, al no disponer de resultados publicados para todos los analizados.

Se comparan los resultados de las secuencias S1, S6, S7, S8 y S10, con los resultados obtenidos en [2], proyecto fin de carrera donde se hace un estudio de distintos algoritmos de segmentación. La métrica utilizada en [2], es igual a la utilizada en este proyecto fin de carrera. Se compara el algoritmo que ha obtenido mejores resultados en [2] para cada secuencia con los cinco sistemas implementados en este proyecto fin de carrera con la configuración k=3 (3 capas máximas para el modelo de fondo). Para ver la comparativa de todas las secuencias en una única tabla, se muestra el valor medio de la medida del parámetro óptimo (definido en la ecuación *eq 7.12*).

Comparación con [2]							
Secuencia	**Método BS en [2]**	**[2]**	**Sistema 1**	**Sistema 2**	**Sistema 3**	**Sistema 4**	**Sistema 5**
S1	Gamma	195,22	186,38	184,54	197,89	197,90	198,16
S6	MoG	153,49	127,32	145,03	167,81	174,08	179,02
S7	MoG	157,31	149,54	148,98	174,54	180,35	181,20
S8	Gamma	192,25	169,00	167,51	192,90	195,42	195,41
S10	Mediana	144,60	127,58	140,89	163,83	181,83	185,47

Tabla 7-49. Resultados comparativos con [2] sobre las secuencias S1, S6, S7, S8, S10

También se han comparado los resultados de este proyecto fin de carrera con los resultados obtenidos por [46] (se dan resultados para un sistema *MoG* y para el sistema propuesto) para las secuencias S5, S11 y S14. La tabla donde se muestra la comparación sigue el mismo formato que la anterior pero en esta nueva tabla se compara utilizando una nueva métrica. (propuesta en [46]) Se trata de la medida de similitud entre dos imágenes binarias. La intersección entre la unión de la máscara de *ground-truth* (*GT*) y la máscara resultado obtenida (*Mask*).

$$S(Mask, GT) = \frac{Mask \cap GT}{Mask \cup GT} \qquad eq\ 7.17$$

Comparación con [46]							
Secuencia	**MoG [46]**	**Propuesto [46]**	**Sistema 1**	**Sistema 2**	**Sistema 3**	**Sistema 4**	**Sistema 5**
S5	0,421	0,706	0,568	0,575	0,132	0,607	0,428
S11	0,445	0,911	0,735	0,678	0,308	0,567	0,568
S14	0,277	0,534	0,174	0,230	0,238	0,419	0,440

Tabla 7-50. Resultados comparativos con [46] sobre las secuencias S5, S11 y S14.

La comparativa con este sistema resulta en estadísticos peores para todos los videos considerados. Esta situación puede deberse a tres factores:

- Las máscaras de *ground-truth* han sido realizadas a mano y no están muy ajustadas a los bordes de los objetos. Esta situación parece contribuir a mejorar los estadísticos de [46].
- El sistema descrito en [46] realiza un tratamiento específico para las sombras, mientras que los propuestos en este proyecto final de carrera no.
- La selección del parámetro de inicialización σ_0 no ha sido la óptima. Para aislar la influencia de este factor se presentan a continuación los valores obtenidos para la secuencia **S11** (*Curtain*) mediante la inicialización óptima de los sistemas propuestos (los valores de esta inicialización resultan de la selección de la β óptima para el análisis obtenida en la sección 7.3.4).

Comparación con [46] (Inicialización óptima)							
Secuencia	**MoG [46]**	**Propuesto [46]**	**Sistema 1**	**Sistema 2**	**Sistema 3**	**Sistema 4**	**Sistema 5**
S11	0,445	0,911	0,735	0,766	0,515	0,623	0,575

Tabla 7-51. Resultados comparativos con [46] sobre la secuencia S11 (inicialización óptima).

7.4 **Problemas solucionados.**

En esta sección se muestra como el sistema final desarrollado en este proyecto fin de carrera (*Sistema 5*) solventa algunos de los problemas del modelado de fondo (definidos en 3.3). Con este propósito se muestra una tabla con resultados (cualitativos) para distintos cuadros del Data-Set utilizado. En a) se muestra el primer cuadro del vídeo, en b) el cuadro evaluado, en c) resultado de la segmentación del *Sistema 5* y en d) el problema asociado al cuadro evaluado.

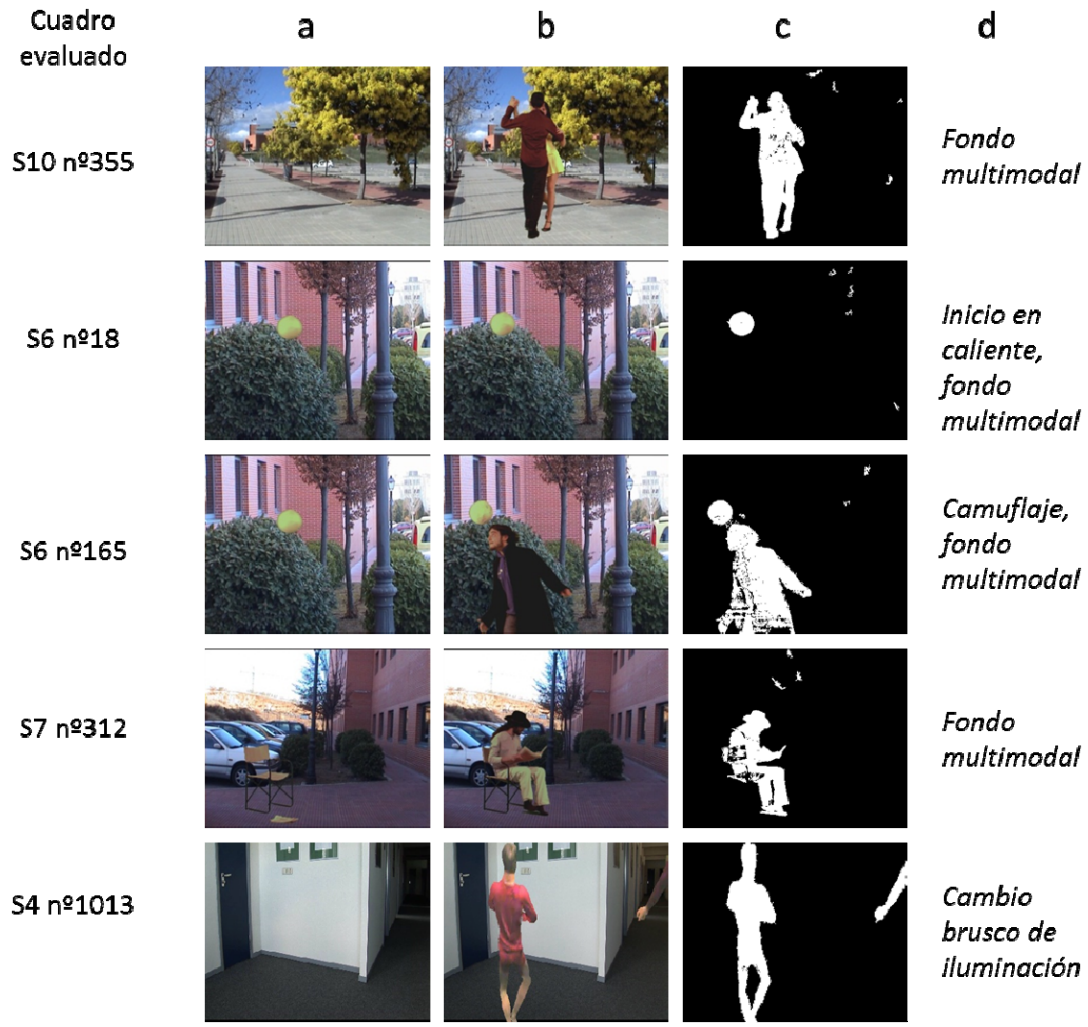


Figura 7-17: Resultados cualitativos y problemas solventados.

7.5 Resultados en proceso de publicación e integración en sistema comercial.

Tras valorar la calidad de los resultados obtenidos, y considerando el interés de la comunidad científica por lo innovador de alguna de las aproximaciones propuestas, se ha decidido junto con el tutor del proyecto y el ponente, redactar un resumen del trabajo realizando un artículo de investigación para su envío a (Electronics Letters), con el título "Class driven Bayesian background modelling" para el cual todavía no se ha recibido respuesta sobre su aceptación. (Ver Anexo C).

Además también se ha estado trabajando de forma paralela en la introducción de este sistema dentro del proyecto RETINAS. Proyecto de investigación en el cual se está trabajando dentro del VPULab para una empresa privada. Los resultados de este proyecto

se adaptan muy bien a los requisitos pedidos pero se detecta un problema en su integración ya que el algoritmo no funciona en tiempo real.

8 Conclusiones y trabajo futuro.

La implementación realizada del sistema *Bayesiano Básico* propuesto en [1] aporta ventajas con respecto a otros algoritmos de segmentación a la hora de segmentar vídeos que contengan fondos multimodales puesto que el modelo de fondo puede soportar varias apariencias distintas para cada pixel (tantas como capas) y la actualización de los parámetros internos asociados a cada capa puede realizarse de manera independiente al resto de las capas.

El sistema incorpora al modelo de fondo todas las apariencias que toma el pixel a lo largo del vídeo sin tener en cuenta que alguna de estas apariencias puede pertenecer a un objeto de frente (píxeles que se quieren separar del fondo). Para evitar la pérdida de píxeles de frente se introduce una distinción en el modelo de fondo en capas de *fondo fiable* y capas de *fondo no fiable*. En las capas de *fondo no fiable* se encuentran píxeles de frente y píxeles de fondo en proceso de modelado.

Este funcionamiento puede verse limitado en determinadas secuencias, aquellas donde los objetos de frente se repiten a lo largo del vídeo o son muy homogéneos. Estas secuencias generan situaciones donde las apariencias de los píxeles de frente ganan la confianza necesaria y se introducen en el *fondo fiable* clasificando erróneamente el frente en los cuadros sucesivos.

La introducción de un umbral de confianza para dividir el fondo en fiable y no fiable, no solo no funciona en determinados vídeos, además necesita un tiempo de inicialización para que nuevas apariencias de fondo lleguen a modelarse en una capa de *fondo fiable*.

Ambas limitaciones se intentan solventar mediante una nueva clasificación del pixel, desarrollada en un primer momento utilizando información de pixel y posteriormente incorporando información de blob. El objetivo principal de esta mejora es la discriminación por un lado de los píxeles de frente y por otro lado de los píxeles de fondo, evitando que las apariencias de frente se modelen en el modelo de fondo.

La clasificación utilizando información de pixel tiene una gran limitación ya que los píxeles de *frente* se 'entremezclan' con los de *fondo estático no modelado*. Debido a ello aunque se consigue la no introducción de píxeles de frente en el modelo de fondo, se introduce una nueva limitación al sistema: se tarda más tiempo en adaptarse a cambios bruscos de iluminación o a *inicios en caliente.* Además, como no se consigue separar los píxeles de frente en una capa independiente, no pueden utilizarse para inicializar y mantener un modelo de frente.

Con la clasificación del pixel utilizando información de blob, se consigue una mejor discriminación entre las diferentes clases: *fondos modelados, fondos dinámicos no modelados, fondos estáticos no modelados y frente*. Gracias a ésto, se consigue la no introducción de píxeles de frente en el modelo de fondo y además se consigue que tanto los *fondos dinámicos no modelados* como los *fondos estáticos no modelados* se modelen en el modelo de fondo.

Adicionalmente, disponer de los píxeles de frente aislados en una nueva capa nos permite cumplir dos objetivos.

- Por un lado se elimina el umbral de confianza del modelo de fondo fusionando las capas de *fondos fiables* y las de *fondos no fiables* en una nueva clase llamada *fondo modelado*. Gracias a esta fusión de clases se reduce el tiempo de inicialización necesario para que una nueva apariencia de fondo participe en el resultado final de la segmentación.
- El segundo objetivo conseguido es que podemos modelar los píxeles de frente en una nueva capa y comparar con ella los píxeles del cuadro de entrada teniendo nueva información para tomar la decisión final en la máscara resultado (frente o fondo).

La clasificación del pixel utilizando información de blob también tiene limitaciones ya que tiene umbrales internos al sistema que aunque se han normalizado en función de la resolución de la secuencia de entrada siguen siendo dependientes del tamaño de los objetos de frente y de otras características de la secuencia de entrada.

El modelado de frente que se ha realizado en este proyecto ha sido un modelo de frente básico donde el modelado se realiza en una única capa. Esta capa tendrá las mismas

características y estará sometida a los mismos procesos de actualización que las utilizadas para el modelo de fondo.

Trabajo Futuro

Una limitación del modelo de frente es que es dependiente del modelo de fondo. El conseguir independizar el modelo de frente del modelo de fondo queda como posible trabajo futuro.

Todos los sistemas propuestos en este proyecto tienen la limitación de que no trabajan en tiempo real y que no son robustos a las sombras de los objetos de frente que pasan por la escena.

Queda como trabajo futuro intentar mejorar la eficiencia del código (ya se ha intentado mejorar la eficiencia con la implementación de mejoras como la descrita en el Anexo B)

Por otro lado se proponen dos estrategias para mejorar el comportamiento del sistema frente a los sombras: algoritmos de post-procesado de las imágenes o introducción de una nueva clase en el sistema (de dedicación específica para el modelado de sombras).

Otra posible mejora para trabajos futuros sería la eliminación de los umbrales internos al sistema o por lo menos el hacerlos más dependientes de las secuencias de entrada. En este trabajo se han normalizado en función de la resolución de la secuencia pero podría ser más robusto el realizar la normalización en función del tamaño del blob.

Finalmente, sería interesante el diseñar un nuevo sistema de segmentación que, basándose en la clasificación de pixel planteada en este trabajo utilice otras características de discriminación distintas a las propuestas, o bien crear una nueva clasificación y crear un modelo Bayesiano que modele cada una de las nuevas clases en capas separadas.

9 **Referencias.**

[1] F. Polikri, O. Tuzel. “Bayesian Background Modeling for Foreground Detection” In: Proceedings of the ACM Visual Surveillance and Sensor Network vol 1(1) pp 55-58, 2005.

[2] S. Herrero, J.Bescós. “Análisis comparativo de técnicas de segmentación de secuencias de vídeo basada en el modelado del fondo” Escuela Politécnica Superior, Universidad Autónoma de Madrid, 2009.

[3] M. Piccardi. “Background substraction techniques: a review” in Proc. of IEEE SMC 2004 International Conference on Systems, Man and Cybernetics, vol 4(1), pp. 3099–3104, 2004.

[4] J.L. Landabaso, M. Pardas. “Cooperative Background Modeling using multiple cameras towards human detection in smart-rooms” In: European Signal Processing Conference EUSIPCO 2006, 2006.

[5] H. Dias, J. Rocha, P. Silva, C. Leao, “Distributed Surveillance System", Conference on Artificial Intelligence, 2005. EPIA 2005, vol 1(1) pp 257-261, 2005.

[6] M. Valera, S. Velastin, “Intelligent Distributed Surveillance Systems: a review”, IEE Proceedings on Vision, Image and Signal Processing, Vol. 152(2) pp 192–204, 2005.

[7] M.M. Chang, M. I. Sezan, A. M. Tekalp, “An algorithm for simultaneous motion estimation and scene segmentation” in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, vol. V(1) pp 221–224, 1994.

[8] R. Lienhart,, C. Kuhmunch, W. Effelsberg. “On the detection and recognition of television commercials”. In: International Conference on Multimedia Computing and Systems 97, vol 1(1) pp. 509–516, 1997.

[9] O. Sukmarg, K.R. Rao. “Fast object detection and segmentation in MPEG compressed domain” In: Proc. IEEE TENCON 2000, vol. 3(1) pp. 364–368, 2000.

[10] K. Toyama, J. Krumm, B. Brumitt, B. Meyers. “Wallflower: Principles and Practice of Background Maintenance” Conference on Computer Vision, IEEE International, vol 1(1) pp 255, 1999.

[11] E.J. Carmona, J. Martínez-Cantos, J. Mirá ‘A new video segmentation method of moving objects based on blob-level knowledge’ Pattern Recognition Letters, vol 29(3) pp 272-285, 2007

[12] C. Stauffer, W.E.L. Grimson, "Adaptive background mixture models for real-time tracking", Computer Vision and Pattern Recognition, 1999. IEEE, vol 2(1) pp 637-663, 1999.

[13] J. Gallego, M. Pardás, J.L. Landabaso. “Segmentation and tracking of static and moving objects in video surveillance scenarios” Conference on Image Processing, ICIP 2008, vol1(1) pp 2716-2719, 2008

[14] W. Lam, C. Pang, N. Yung. “Highly accurate texture-based vehicle segmentation method“ SPIE International Society for Optical Engineering, vol 43(3) pp 591-603, 2004.

[15] S.C. Cheung, C. Kamath. “Robust background subtraction with foreground validation for urban traffic video” EURASIP J. Appl. Signal Process, vol 2005(1) pp 2330-2340, 2005.

[16] O. Javed, K. Shafique, M. Shah. “A hierarchical approach to robust background subtraction using color and gradient information”. Motion and Video Computing, 2002. Procceedings. Workshop on, vol 1(1) pp 22-27, 2002.

[17] M. Cristani, M. Bicego, V. Murino. “Multi-level background initialization using Hidden Markov Models” In First ACM SIGMM International workshop on Video surveillance, pp 11-20, 2003.

[18] B. Gloyer, H.K. Aghajan, K. Siu, “Video-based freeway monitoring system using recursive vehicle tracking” In Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol 2421(1) pp 173-180, 1995.

[19] J. Wu, C. Gu. "The Design and Implementation of Real-Time Automatic Vehicle Detection and Counting System" International Conference on Information Engineering and Computer Science, 2009. ICIECS 2009, vol 1(1) pp.1-4, 2009.

[20] R. Ewerth, B. Freisleben. “Frame difference normalization: an approach to reduce error rates of cut detection algorithms for MPEG videos”. Conference on Image Processing, 2003. ICIP 2003, vol 2(1) pp 1009-1012, 2003.

[21] D. Koller, J. Weber, T. Huang, J. Malik. “Toward robust automatic traffic scene analysis in realtime” International conference Pattern Recognition 1994, vol 1(1) pp 126-131, 1994.

[22] C.R. Wren, A. Azarbayejani, T. Darrell, A.P. Pentland. “Pfinder: Real-time tracking of the human body” IEEE Transactions on Pattern Analisis and Machine Intelligence, vol 19(7) pp 780-785, 1997.

[23] D. Toth, T. Aach, V. Metzler. “Bayesian spatio-temporal motion detection under varying illumination”. In Proc. of EUSIPCO, pp 2081-2084, 2000.

[24] Dar-Shyang Lee, J.J. Hull, B. Erol. “A Bayesian framework for Gaussian mixture background modeling”. Conference on Image Processing, ICIP 2003, vol 3(1) pp 973-6, 2003.

[25] J. Gallego, M. Pardás. “Bayesian foreground segmentation and tracking using pixel-wise background model and region based foreground model” Conference on Image Processing, ICIP 2009, vol 1(1) pp 3205-3208, 2009.

[26] G. Gordon, T. Darrell, M. Harville, J. Woodfill. “Background estimation and removal based on range and color” CVPR, 1999. IEEE Comp Society Conf. on Comp Vision and PattRecog (CVPR'99) vol 2(1) pp 637-663.

[27] C. Stauffer, W. Grimson. “Learning patterns of activity using real time tracking”. In IEEE Transactions. on Pattern Analysis and Machine Intelligence, vol 22(8) pp 747-757, 2000.

[28] M. Harville, G. Gordon, J. Woodfill. “Foreground segmentation using adaptive mixture models in color and depth” IEEE Workshop on Detection and Recognition of Events in Video, 2001, vol 1(1) pp 3-11, 2001.

[29] A. Elgammal, R. Duraiswami, D. Harwood, L.S. Davis. “Background and foreground modeling using nonparametric kernel density estimation for visual surveillance” Proceedings of the IEEE, vol 90(7) pp 1151-1163, 2002.

[30] O. Feron and A. Mohammad-Djafari, “A hidden Markov model for image fusion and their joint segmentation in medical image computing”, MICCAI, St. Malo, France, 2004.

[31] C. Benedek, T. Sziranyi. “Bayesian Foreground and Shadow Detection in Uncertain Frame Rate Surveillance Videos” Transactions on Image Processing, IEEE, vol 17(4) pp 608-621, 2008.

[32] Y. Sheikh, M. Shah. “Bayesian Modeling of Dynamic Scenes object Detection” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 27(1) pp 1778-1792, 2005.

[33] P. Brault, A. Mohammad-Djafari. “Bayesian segmentation and motion estimation in video sequences using a Markov-Potts model” Proceedings of the 5th WSEAS International Conference on Applied Mathematics pp 11:1--11:6, 2004.

[34] R. Boesch, Z. Wang. “Segmentation Optimization for aerial images with spatial constraints“ The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol 37(B4). 2008.

[35] J. Chen, T.N Pappas “Adaptive image segmentation based on color and texture” Conference on Image Processing. 2002. Proceedings, vol 3(1) 777-780, 2002

[36] F. Porikli, O.Tuzel. “Covariance tracking Using Model Update Based On Riemannian Manifolds”, Computer Vision and Pattern Recognition, 2006.

[37] W. Förstner, B. Moonen. “A metric for covariance matrices” .Technical report Department of Geodesy and Geoinformatics, 1999.

[38] E. López-Rubio, R. Luque-Baena. “Stochastic approximation for background modelling”. Computer Vision and Image Understanding vol 115(6) pp 735-749, 2011.

[39] Y. Weiss, E.H. Adelson. “Motion estimation and segmentation using a recurrent mixture of experts architecture” IEEE Workshop vol 1(1) pp293-302, 1995.

[40] C. M. Bishop, M. Svensen. “Bayesian Hierarchical Mixtures of Experts” In U. Kjaerulff and C. Meek, editors, Proceedings Nineteenth Conference on Uncertainty in Artificial Intelligence, pages 57–64. Morgan Kaufmann, 2003.

[41] M. Cristani, M. Farenzena, D. Bloisi, V. Murino. “Background subtraction for automated multisensor surveillance: a comprehensive review”. EURASIP J. Adv. Signal Process, vol 2010() pp 43:1--43:24, 2010.

[42] G. Welch and G. Bishop. “An Introduction to the Kalman Filter” Technical Report 95–041. University of North Carolina at Chapel Hill, 2001.

[43] X. Zhang, J.Yang, “Foreground segmentation based on selective foreground model”. Electronics Letters, vol 44(14) pp 851-852, 2008.

[44] F. Tiburzi, M. Escudero, J. Bescós, J.M. Sanchez. “A ground truth for motion-based video-object segmentation”. In Proceedings: International Conference on Image Processing (ICIP 08). Vol 1(1) pp 17-20, 2008.

[45] VSSN 2006 Call for Algorithm Competition in Foreground/Background Segmentation (2006), http://mmc36.informatik.uni-augsburg.de/VSSN06_OSAC/

[46] L. Li, W. Huang, I. Y.-H. Gu, and Q. Tian. “Statistical modelling of complex backgrounds for foreground object detection” IEEE Transactions on Image Processing, vol 13 (11), 2004.

Anexos.

Anexo A: desarrollo del modelado de cada capa.

Los sistemas de segmentación que se han implementado en este proyecto fin de carrera se componen de un conjunto de capas donde se van modelando los distintos modos que aparecen en los vídeos. Por tanto dependiendo de la multimodalidad del fondo que caracteriza el escenario de análisis, el sistema se inicializará con un número mayor o menor de capas. Siempre se suele inicializar con un número de capas mayor al necesario. Generalmente entre tres y cinco pero depende del vídeo analizado. Cuando inicializamos el modelo, se inicializa una de las capas a la primera imagen del vídeo y el resto de capas se dejan vacías. El sistema y por tanto el modelado de cada capa está basado en el paper [1] de Fatih Porikli.

En cada capa se asume una distribución normal con media M y covarianza Σ. La media y la varianza se desconocen y se modelan como variables aleatorias. Usando el teorema de Bayes, la densidad de probabilidad conjunta a posteriori puede escribirse como:

$$P(\mathrm{M},\Sigma \mid X) \alpha\, p(X \mid \mathrm{M},\Sigma)\, p(\mathrm{M},\Sigma) \qquad \textit{eq A.1}$$

Para conseguir una estimación recursiva bayesiana con una nueva observación, la densidad de probabilidad conjunta a priori $P(\mathrm{M},\Sigma)$ debe tener la misma forma que la probabilidad conjunta a posteriori $P(\mathrm{M},\Sigma \mid X)$. Condicionando a la varianza, la densidad de probabilidad conjunta a priori se escribe:

$$P(\mathrm{M},\Sigma) = p(\mathrm{M} \mid \Sigma)\, p(\Sigma) \qquad \textit{eq A.2}$$

Estas condiciones se cumplen por ejemplo si asumimos una distribución inversa de Wishart para la covarianza y una distribución normal de variable aleatoria para la media. La distribución inversa de Wishart es una generalización de la distribución inversa-X^2. La parametrización queda:

$$\begin{aligned} &\Sigma \sim Inv-Wishart_{v_{t-1}}\left(\Lambda_{t-1}^{-1}\right) \\ &M \mid \Sigma \sim N\left(m_{t-1}, \Sigma / \kappa_{t-1}\right) \end{aligned} \qquad \textit{eq A.3}$$

Donde v_{t-1} y Λ_{t-1} son los grados de libertad y la matriz de escala para la distribución inversa de Wishart, m_{t-1} es la media a priori y κ_{t-1} es el número de muestras utilizadas para el modelo a priori. Con estas asunciones, la densidad de probabilidad conjunta a priori tomando un espacio de características en tres dimensiones es:

$$P(m,\Sigma) \propto |\Sigma|^{((v_{t-1}+3)/2+1)} \; xe^{\left(-\frac{1}{2}tr\left(\Lambda_{t-1}\Sigma^{-1}\right)-\frac{\kappa_{t-1}}{2}(M-m_{t-1})^T\Sigma^{-1}(M-m_{t-1})\right)} \qquad \textit{eq A.4}$$

Modelo de adaptación.

Cuando un píxel pertenece a una de las capas de fondo ya inicializadas (nueva observación), se actualiza el modelo siguiendo las siguientes ecuaciones:

$$\begin{aligned} &v_t = v_{t-1}+1 \;,\; \kappa_t = \kappa_{t-1}+1 \\ &m_t = m_{t-1}\frac{\kappa_{t-1}}{\kappa_{t-1}+1} + x_t\frac{1}{\kappa_{t-1}+1} \\ &\Lambda_t = \Lambda_{t-1} + \frac{\kappa_{t-1}}{\kappa_{t-1}+1}(x_t-m_{t-1})(x_t-m_{t-1})^T \end{aligned} \qquad \textit{eq A.5}$$

Los nuevos parámetros combinan la información a priori del modelo con la información de la nueva muestra observada. La media a posteriori calculada m_t es un promedio ponderado entre la media a priori m_{t-1} y la observación x_t.

De la misma manera cuando se detecta que la muestra de entrada no pertenece a alguno de los modos inicializados en el modelo, se actualiza la capa con la siguiente ecuación:

$$\begin{aligned} &\kappa_t = \kappa_{t-1}-1 \\ &Con \longrightarrow \kappa_t \geq 10 \end{aligned} \qquad \textit{eq A.6}$$

Con esta fórmula se baja la fiabilidad de la capa y por tanto la probabilidad de que las nuevas muestras pertenezcan a este modo ya que la última muestra que ha entrado al modelo no pertenecía a esta capa. Se destaca que el número de muestras utilizadas para el modelo a priori nunca puede ser menor de diez.

Modelo de inicialización.

Las ecuaciones para inicializar el modelo de un píxel dentro de una de las capas es la siguiente:

$$\begin{aligned} v_0 &= 10 \\ \kappa_0 &= 10 \\ m_0 &= x_0 \\ \Lambda_0 &= (v_0 - 4)16^2 I \end{aligned} \qquad \textit{eq A.7}$$

Donde se define I como la matriz de identidad de tres dimensiones: $I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

Integrando la densidad de probabilidad conjunta a posteriori con respecto a Σ se consigue la densidad marginal de probabilidad a posteriori para la media:

$$p(M \mid X) \propto t_{v_{t-2}}\left(M \mid m_t, \Lambda_t / (\kappa_t (v_t - 2))\right) \qquad \textit{eq A.8}$$

Donde $t_{v_{t-2}}$ es una t-distribución con $v_t - 2$ grados de libertad. Calculando la esperanza de la media y la covarianza en el instante *t* para la distribución marginal a posteriori se tiene:

$$M_t = E[M \mid X] = m_t \qquad \textit{eq A.9}$$

$$\Sigma_t = E[\Sigma \mid X] = (v_t - 4)^{-1} \Lambda_t \qquad \textit{eq A.10}$$

La medida de confianza de la capa C que mide la fiabilidad del modelo es igual a uno entre el determinante de Σ de $M \mid X$.

$$C = \frac{1}{|\Sigma_{M|X}|} = \frac{\kappa_t^3 (v_t - 2)^4}{(v_t - 4)|\Lambda_t|} \qquad \textit{eq A.11}$$

Hay que destacar que si la media marginal a posteriori tiene una varianza alta, el modelo será poco fiable.

Cuanto más se parece la muestra al modelo, más aumenta la confianza, ésta es una de las mayores ventajas del modelado Bayesiano, puesto que se tiene una estimación de como de bueno es el modelo.

Anexo B: cálculo de la inversa de la matriz de covarianza completa.

Se ha trabajado en cálculo matricial para que el coste computacional de introducir la matriz de covarianza completa en el cálculo de la distancia de Mahalanobis 3.5.5.1 sea el mínimo posible. Se define la matriz de confianza como:

$$\Sigma = \begin{pmatrix} a & b & c \\ b & d & e \\ c & e & f \end{pmatrix} \qquad \textit{eq B.1}$$

Donde cada letra indica la posición dentro de la matriz.

Las ecuaciones utilizadas para el cálculo del determinante de la matriz de confianza y de su inversa son las siguientes:

$$\det\begin{pmatrix} a & b & c \\ b & d & e \\ c & e & f \end{pmatrix} = adf - c^2 d + 2bce - e^2 a - b^2 f \qquad \textit{eq B.2}$$

$$\begin{pmatrix} a & b & c \\ b & d & e \\ c & e & f \end{pmatrix}^{-1} = \left(\det\begin{pmatrix} a & b & c \\ b & d & e \\ c & e & f \end{pmatrix}\right)^{-1} \begin{pmatrix} a' & b' & c' \\ b' & d' & e' \\ c' & e' & f' \end{pmatrix} \qquad \textit{eq B.3}$$

$$\begin{aligned} a' &= df - e^2 \\ b' &= ce - bf \\ c' &= be - cd \\ d' &= af - c^2 \\ e' &= bc - ae \\ f' &= ad - b^2 \end{aligned} \qquad \textit{eq B.4}$$

Las ecuaciones indicadas también han sido utilizadas en trabajos como en [38] donde se indica que con ellas se puede trabajar en tiempo real.

Anexo C: publicaciones.

Título: Class driven Bayesian background modelling.

Autores: Alfonso Colmenarejo, Marcos Escudero-Viñolo, Jesús Bescós.

Revista: *Electronics Letters*.

Estado: Enviado artículo y esperando respuesta.

PRESUPUESTO.

1) Ejecución Material

- Compra de ordenador personal (Software incluido)..... ..... 2.000 €
- Alquiler de impresora láser durante 6 meses ..... 50 €
- Material de oficina ..... 150 €
- Total de ejecución material ..... 2.200 €

2) Gastos generales

- sobre Ejecución Material ..... 352 €

3) Beneficio Industrial

- sobre Ejecución Material ..... 132 €

4) Honorarios Proyecto

- 840 horas a 15 € / hora..... 12600 €

5) Material fungible

- Gastos de impresión..... 80 €
- Encuadernación..... 200 €

6) Subtotal del presupuesto

- Subtotal Presupuesto..... 15080 €

7) I.V.A. aplicable

- 18% Subtotal Presupuesto ..... 2714,4 €

8) Total presupuesto

- Total Presupuesto..... 17794,4 €

Madrid, Julio de 2011

El Ingeniero Jefe de Proyecto

Fdo. Alfonso Colmenarejo Rubio.
Ingeniero Superior de Telecomunicación.

PLIEGO DE CONDICIONES

Este documento contiene las condiciones legales que guiarán la realización, en este proyecto, de un sistema de segmentación de secuencias de vídeo basado en modelado del fondo mediante capas desarrollado en este PFC. 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 partida alzada 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.

Class driven Bayesian background modelling

Journal:	*Electronics Letters*
Manuscript ID:	Draft
Manuscript Type:	Letter
Date Submitted by the Author:	n/a
Complete List of Authors:	Colmenarejo, Alfonso; Universidad Autónoma de Madrid,, TEC Escudero-Viñolo, Marcos; Universidad Autónoma de Madrid, TEC Bescós, Jesús; Universidad Autónoma de Madrid, TEC
Keywords:	Multilayer background modelling, Bayesian background modelling, Pixel clasification, Foreground, dynamic backgrounds, static backgrounds, illumination changes, hot starts, camouflage

Class driven Bayesian background modelling

Alfonso Colmenarejo, Marcos Escudero-Viñolo, Jesús Bescós

A background subtraction video segmentation algorithm that works by modelling the different appearances of a pixel in a set of independent layers is proposed. The main contribution of this work to the existing approaches is the use of an a priori classification scheme that classifies the pixel before the model updating. By means of this classification, the algorithm is capable to isolate the pixel foreground samples and to avoid its influence in the updating and discrimination processes of the subsequent frames. Obtained results demonstrate the adequate performance of the algorithm in the presence of highly dynamic backgrounds, foreground-background similarity, hot starts and abrupt illumination changes.

Introduction: There are several strategies to model dynamic backgrounds in background subtraction algorithms. The most popular is to describe the pixel evolution by a parametric model resulting of a combination of simpler sub models (as the Gaussians in a Mixture of Gaussians). These strategies are capable of handling several modes, one per sub model [1]. However, the updating of each sub-model affects the others. Alternatively, some authors plead for the representation of the background model in k layers. Following a similar approximation to [2], the likelihood of a new sample belonging to a background layer stands:

$$p\left(BG_t, z_t \mid x_t\right) = p\left(BG_t \mid z_t, x_t\right) p\left(z_t \mid x_t\right) \quad (1)$$

The main advantages of using multilayer schemes in the updating and discrimination stages of a background subtraction algorithm are: I) Modifications of the intra-layer models do not affect the rest of the layers. II) The likelihood of each sample belonging to a layer and each layer to the background are isolated (see equation (1)).

State of the Art systems in the area of multilayer background modelling use Bayesian based schemes at the model update processes. Two different designs can be differentiated: one option is to have several layers modelling one class, as in [3] where class assignment is performed by thresholding of the reliability of each layer being the class modelled. In this strategy, the thresholding stage is crucial as misclassifications would be propagated in the model. An alternative is to model one class at each layer (background, foreground, etc.) [4], here the matching process between each new sample and the layer is the key factor, as it determines the class. In this work, a hybrid strategy that combines both schemes is presented. Proposed layered scheme avoids propagation of wrongly classified pixels in the model.

Pixel Classification: A pixel, understood as a temporal volume placed on a fixed coordinate at each frame, can be assigned to different classes along the video. There are five possible classes for a pixel sample: modelled background (MBG), un-modelled dynamic background (UDBG), un-modelled static background (USBG), foreground (FG) and unclassified (U). MBG pixel samples are those which appearance has been previously modelled in the background model. UDBG pixels have not been previously modelled, but its previous and posterior samples are classified as MBG. USBG pixel samples present un-modelled appearances are prelude of equal appearances in subsequent samples and are not followed by MBG samples. FG pixel samples are either defined as un-modelled in the background model and followed by unequal appearances or as previously modelled in the foreground model. Finally U pixels samples are potential samples of every other class, we stored them in an intermediate layer for further analysis.

Classifying the pixel: Every pixel sample of a new frame is a U pixel. The aim is to minimize both the number of U and misclassified samples. As described in Figure 1, the system follows a hierarchical strategy. Next sections describe its modules.

Background model: A multilayer scheme inspired in [3] is proposed. Each layer models an appearance of each pixel, and then background multimodality is considered. A confidence measure is assigned to each layer in order to describe its likelihood of being background. However, differently than in [3], this confidence is not used to distinguish between reliable and unreliable background. Due to the proposed updating scheme, every appearance modelled can be considered reliable background, thus for MBG samples: $p\left(BG_t \mid z_t, x_t\right)=1$. Nevertheless, layer confidence would be used to evaluate the temporal evolution of a pixel. Confidence is straight proportional to the number of observed samples that match at each layer and inverse proportional to the distance among the matched samples.

For evaluation of the new samples, the layers are first ordered according to its confidence value. Intra-layer matching ($p\left(z_t \mid x_t\right)$) is performed by a full covariance Mahalanobis test (M). However, in order to avoid empirical thresholding of the distance and to adapt to the appearance characteristics, we propose to model the distance's evolution by a single Gaussian at each layer. Then, pixel samples with similar appearance to the modelled would result in expected distances (falling inside the layer Gaussian). Distance Gaussians would be updated by a running-average scheme with an envelope shape (the updating factor that weights the influence of new samples) that varies with the layer confidence: first, low confidences indicates that the appearance is under modelling and the Gaussian does not reliable represent the distance evolution then, to minimize the influence of outliers the influence of new samples is chosen to be low. When the confidence increases, the updating factor does the same in order to adapt the model to the progressive variations of the appearances. Finally, to avoid over-training, when a layer has reached a high confidence, updating factor returns to the initial value and starts growing again.

Foreground model: The operation at the foreground layer is similar to that performed at one layer of the background model. There are three main differences: first, the model is fed with FG samples. Second, the model is translated before the matching process and third, a new strategy to eliminate old FG samples is designed. The foreground translation is performed by a simple but efficient *Kalman* filter between the FG samples in the model and the U samples coming out from the background model. After translation, the likelihood of being FG of every sample in the incoming frame is evaluated. Matched samples are classified as FG even if they have been previously classified as MBG. This strategy makes the foreground model dependant of the output of the background model. However, it achieves adequate reclassification of camouflaged pixels especially in the presence of homogeneous foregrounds. Finally, FG samples confidences are used to eliminate old samples from the model. If the confidence value of a sample decreases continuously, the sample is deleted from the model.

UDBG detection: After background and foreground model comparison, confidence evolution of the remaining samples is evaluated (C). Oscillations in a MBG pixel confidence value (that is alternative increasing and decreasing stages) are consequence of alternative periods of new samples matching and un-matching. This in turn, is an indicator of classical behaviour of a multi-modal pixel. Then UDBG pixels are detected by temporal evaluation of the confidence evolution of its corresponding MBG pixels in an analysis window. If the confidence of every modelled appearance does not descend continuously, but varies, the sample is classified as UDBG and would be used to initialize a new appearance in the layer model. The rest of the samples feed the next module.

USBG and FG discrimination: Remaining pixel samples include USBG, FG and U samples. To discriminate among them two blob based descriptors are computed.

Blobs B_t are extracted from the set of remaining samples and the intermediate layer (B_{t-n}) and compared via a *Kalman* filter. Being (x_t, y_t) the coordinates of its mass centre, the descriptors are (O & M):

$$Overlapping = \frac{B_t \cap B_{t-n}}{B_t \cup B_{t-n}}, \quad Motion = \sqrt{(x_{t-n} - x_t)^2 + (y_{t-n} - y_t)^2} \tag{2}$$

USBG pixels are those that belongs to a blob with an $Overlapping = 1$ and a $Motion \approx 0$. On the other side, new FG samples are detected when $Overlapping > 0.5$ and $Motion > th$, where th has been empirically selected to be equal to the 5% of the frame diagonal. The rest of the samples are stored in the intermediate layer and used at the next frame discrimination.

Experimental results: Final segmentation masks include FG and U pixels at each frame. Results are obtained via evaluation of the dataset described at [5]. Videos are selected due to its highly dynamic background, repetitive foreground and shadows absence. Metrics used for comparison are the $FScore_1$ (FS1) and the $FScore_0$ (FS0) as described in [6]. Quantitative comparison (Figure 2) is performed against classical state of the art methods (mono-layers supporting multimodality: [1], [7], [8] and multilayer Bayesian [3]). Qualitative results of the system performance including common background problems are included at Figure 3, where a common video used to evaluate segmentation performances is also included.

Discussion and conclusions: Results indicate that proposed method achieves adequate performance in the presence of illumination changes, hot starts, camouflage situations and highly dynamic backgrounds. Additionally, it does not propagate misclassifications and avoid the introduction of foreground in the model (differently than [3]). Further research must include the introduction of a cast shadows devoted layer.

References

1 STAUFFER, C., and GRIMSON, W. : 'Learning patterns of activity using real time tracking'. In IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, Vol 22(8), pp 747-757.

2 BRAULT, P., and MOHAMMAD-DJAFARI, A. : 'Bayesian segmentation and motion estimation in video sequences using a Markov-Potts model'. Proceedings of the 5th WSEAS International Conference on Applied Mathematics, 2004, Vol 1(1), pp 11:1--11:6.

3 PORIKLI, F., and TUZEL, O. : 'Bayesian Background Modeling for Foreground Detection'. In: Proceedings of the ACM Visual Surveillance and Sensor Network, 2005, Vol 1(1), pp 55-58.

4 BENEDEK, C., and SCIRANYI, T. : 'Bayesian Foreground and Shadow Detection in Uncertain Frame Rate Surveillance Videos'. Transactions on Image Processing, IEEE, 2008, Vol 17(4), pp 608-621.

5 TIBURZI, F., ESCUDERO, M., BESCÓS, J., and SÁNCHEZ, J.M. : 'A ground truth for motion-based video-object segmentation'. In Proceedings: International Conference on Image Processing (ICIP 08), 2008, Vol 1(1), pp 17-20.

6 HERRERO, S., and BESCÓS, J. : 'Background Subtraction Techniques: Systematic Evaluation and Comparative Analysis' . In Proceedings of Advanced Concepts for Intelligent Vision Systems, 11th International Conference, ACIVS 2009, 2009, Vol. 1 (1), pp.33-42.

7 ELGAMMAL, A., DURAISWANI, R., HARWOOD, D., and DAVIS, L.S. : 'Background and foreground modeling using nonparametric kernel density estimation for visual surveillance' Proceedings of the IEEE, 2002, Vol 90(7), pp 1151-1163.

8 CAVALLARO, A., STEIGER, O., EBRAHIMI, T. : 'Semantic video analysis for adaptive content delivery and automatic description'. IEEE Transactions on Circuits and Systems for Video Technology, 2005, Vol 15(10), pp 1200–1209.

Authors' affiliations:

A.Colmenarejo, M. Escudero-Viñolo (corresponding author), J.Bescós
(Video Processing and Understanding Lab, Escuela Politécnica Superior, Universidad Autónoma de Madrid, Ciudad Universitaria de Cantoblanco, 28049 Madrid, Spain)

E-mail: {alfonso.colmenarejo, marcos.escudero, j.bescos}@uam.es

Figure captions:

Figure 1. System overview.

Figure.2. Comparative quantitative results.

Figure.3. Qualitative results and system performance a) First frame of the video, b) Evaluated frame {355, 18, 160, 312, 1013}, c) Segmentation mask, d) Associated background problems.

Figure 1

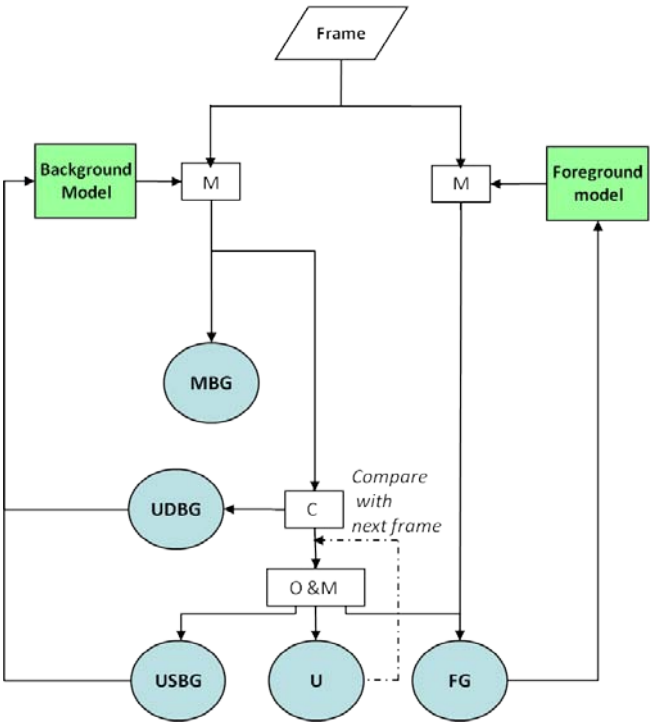

Frame
Background Model
M
M
Foreground model
MBG
UDBG
C
Compare with next frame
O &M
USBG
U
FG

Figure 2

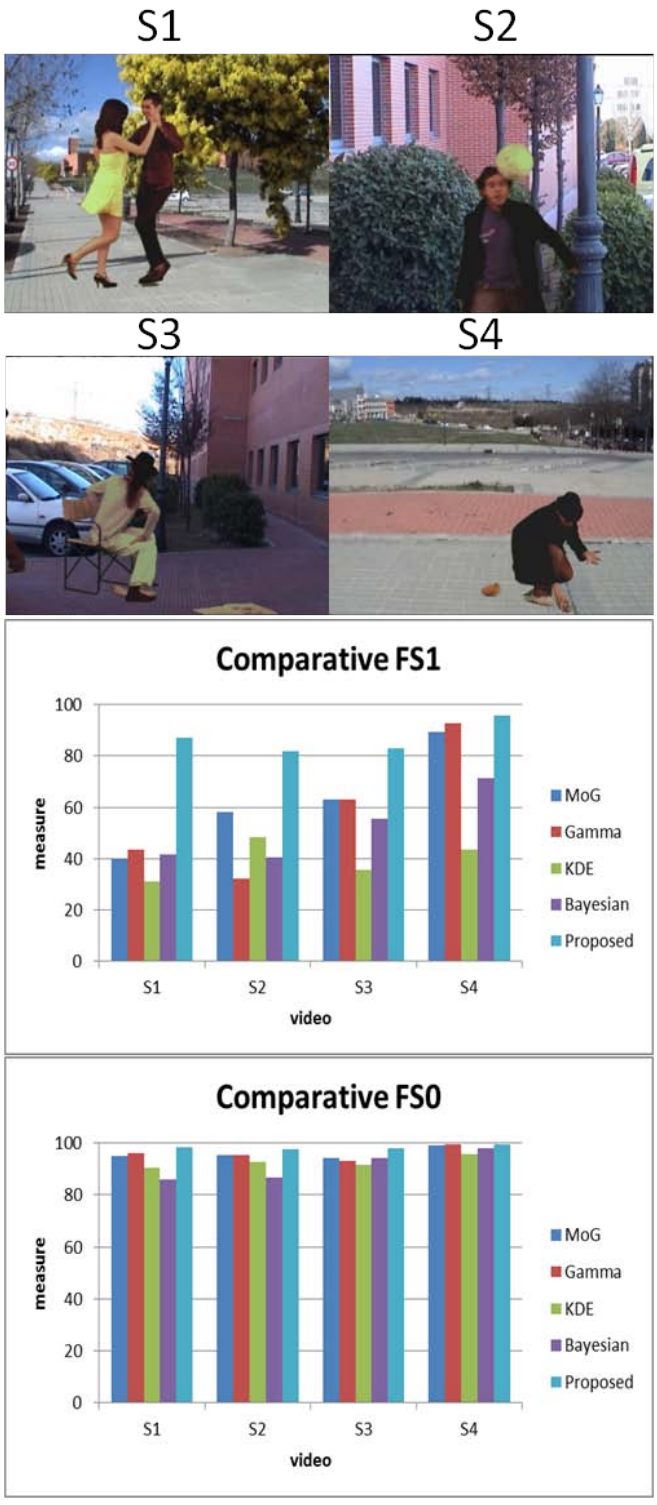
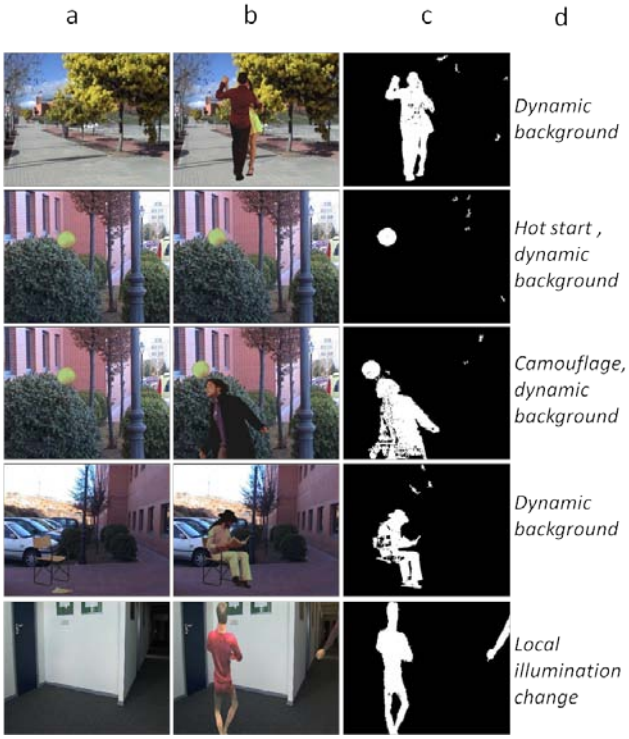
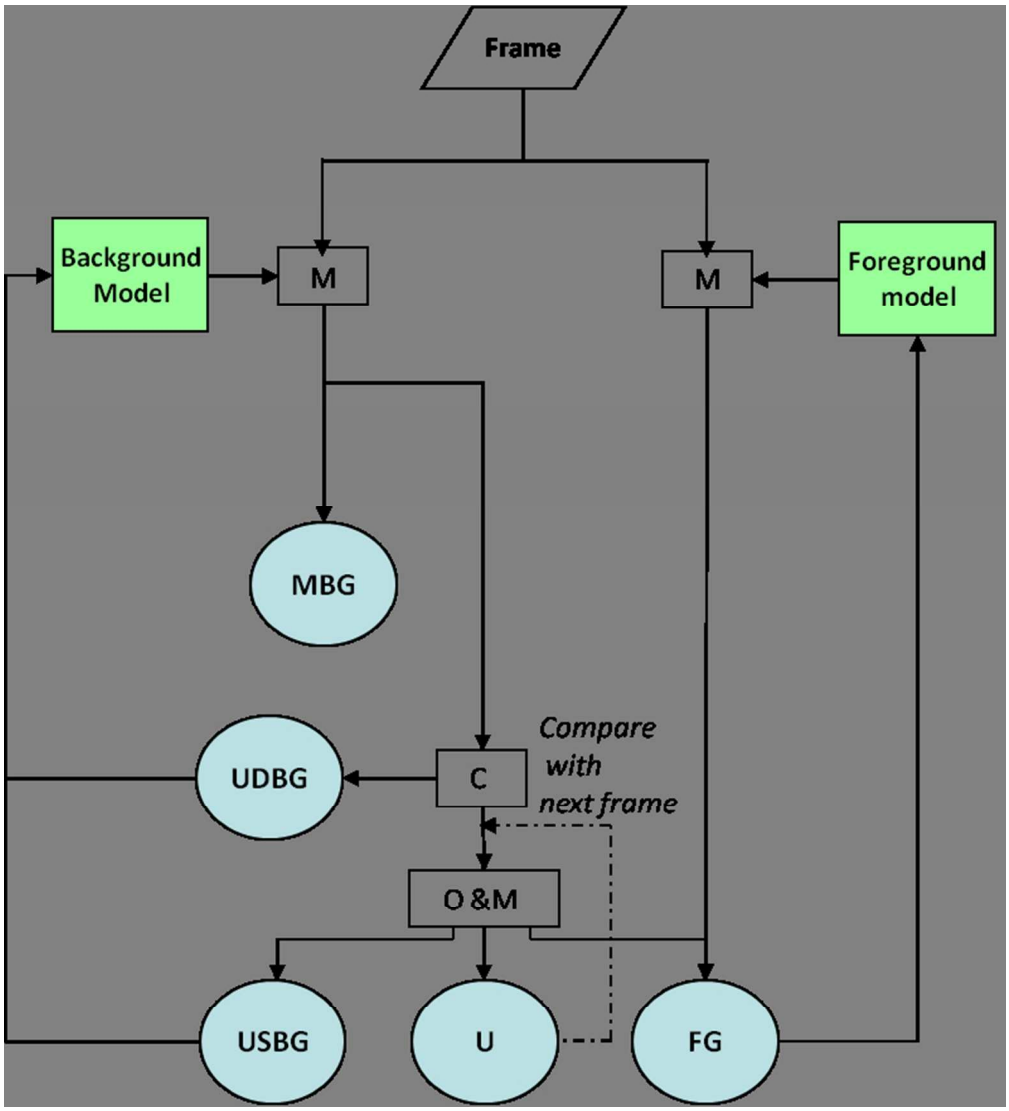
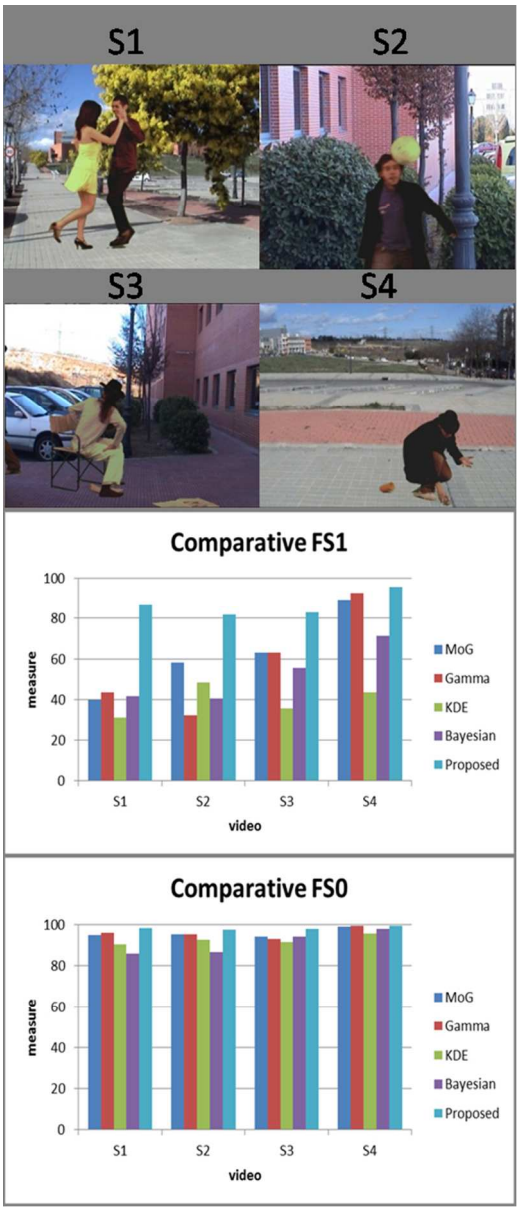

S1
S2
S3
S4
Comparative FS1
100
80
60
40
20
0
measure
S1
S2
S3
S4
video
MoG
Gamma
KDE
Bayesian
Proposed
Comparative FS0
100
80
60
40
20
0
measure
S1
S2
S3
S4
video
MoG
Gamma
KDE
Bayesian
Proposed

Figure 3

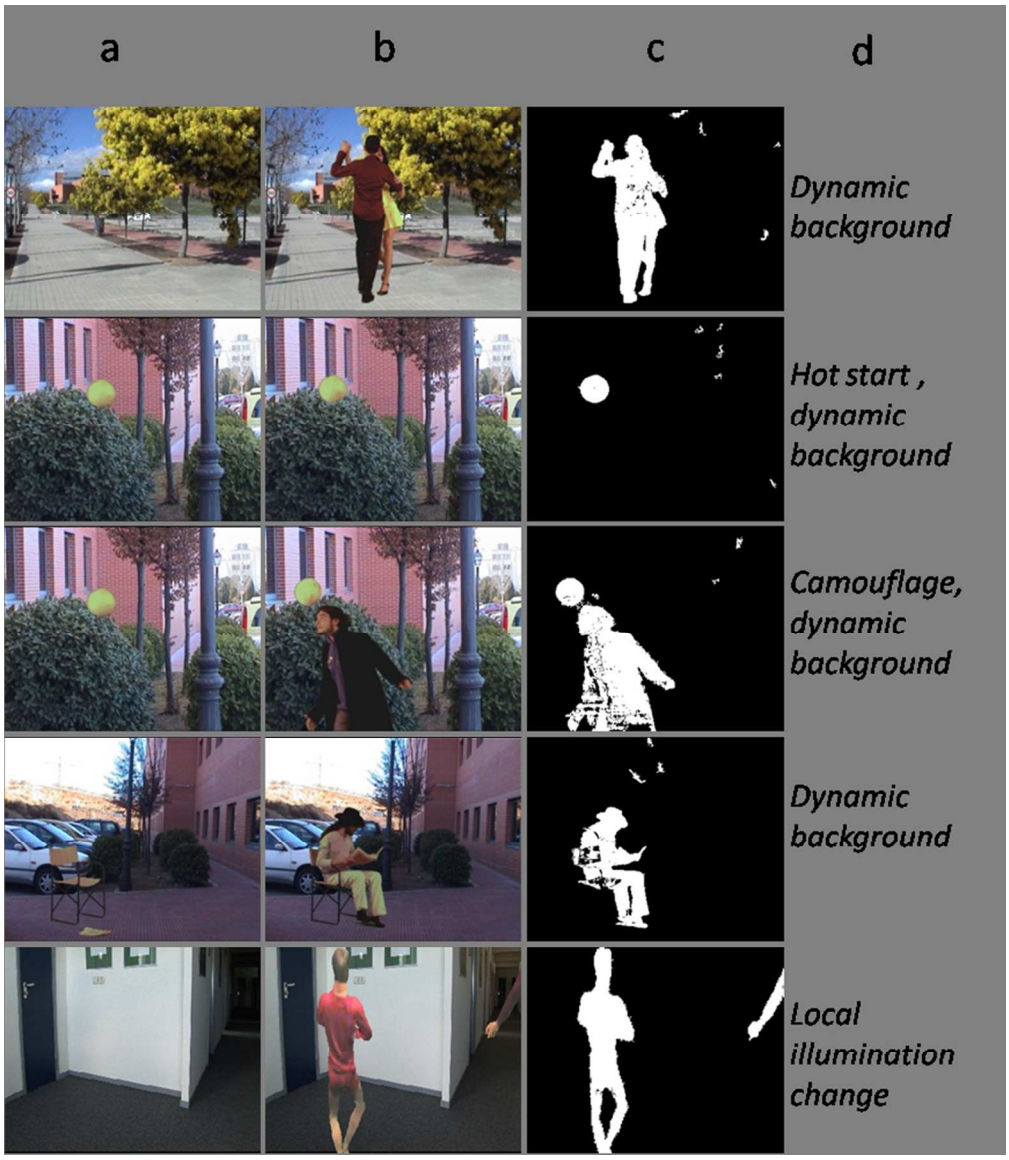

a
b
c
d
Dynamic background
Hot start , dynamic background
Camouflage, dynamic background
Dynamic background
Local illumination change



System overview.



Comparative quantitative results.



Qualitative results and system performance a) First frame of the video, b) Evaluated frame {355, 18, 160, 312, 1013}, c) Segmentation mask, d) Associated background problems.