

UNIVERSIDAD AUTÓNOMA DE MADRID

ESCUELA POLITÉCNICA SUPERIOR



PROYECTO FIN DE CARRERA

**Integración de información de movimiento en la
segmentación de secuencias de vídeo basada en el
modelado de fondo**

Pablo Cervera Benito

Julio 2010

Integración de información de movimiento en la segmentación de secuencias de vídeo basada en el modelado de fondo

AUTOR: Pablo Cervera Benito
TUTOR: Jesús Bescós Cano



Video Processing and Understanding Lab
Dpto. de Ingeniería Informática
Escuela Politécnica Superior
Universidad Autónoma de Madrid
Julio de 2010

Palabras clave

Segmentación, vídeo, modelado de fondo, sustracción de fondo, objetos en movimiento, frente, fondo, FG ('Foreground'), BG ('Background'), Gaussiana Simple (SG), estimación de movimiento, flujo óptico, Block Matching, vector de movimiento, clustering, ancho de banda, precisión, 'recall', 'fscore'.

Resumen

El principal objetivo de este PFC es la implementación de un algoritmo que sea capaz de realizar una integración de la información del movimiento de los objetos de una escena en la segmentación de secuencias de video basada en el modelado de fondo. Para ello se ha seleccionado un método de segmentación que será mejorado gracias al estudio de la información de movimiento obtenida mediante la estimación del movimiento de las secuencias de video. Para la estimación se han estudiado exhaustivamente los métodos basados en el flujo óptico y *Block Matching* y posteriormente se ha realizado un estudio de las regiones de movimiento de las secuencias. Como resultado se ha obtenido un método nuevo que mejora la segmentación por medio del estudio de la coherencia del movimiento de los objetos/personas de la imagen.

Agradecimientos

En primer lugar, quiero agradecer a mi tutor, Jesús Bescós, por permitir que realizase este Proyecto Fin de Carrera en el VPULab, y por su apoyo y dedicación durante todos estos años de estudios.

Gracias a mi familia por sus consejos y todo el amor y el apoyo que me han dado durante toda mi vida, por los valores que me han inculcado y por hacer de mí una persona principalmente feliz.

Gracias a mis compañeros de universidad por haber sido un pilar fundamental a la hora de realizar mis estudios. Porque sin ellos hubiera sido muy difícil. Gracias a Fabio, Luis, Ángel, María, María Riesco, Juanjo, Oscar, Diego por todos los momentos de risas, de estudio, de prácticas, de viajes, de abrazos y de mucho cariño, siempre os recordaré.

Gracias también a todos los profesores con los que he coincidido en estos años, su conocimiento ha servido para instruirme y han fomentado el interés por aprender que sigo teniendo cada día de mi vida.

Muchas gracias.

***Pablo Cervera,
Julio 2010.***

Índice de contenidos

1.		
INTRODUCCIÓN.....		1
1.1.	MOTIVACIÓN.....	1
1.2.	OBJETIVOS.....	3
1.3.	ORGANIZACIÓN DE LA MEMORIA.....	6
2.....		7
ANTECEDENTES.....		7
2.1.	SEGMENTACIÓN DE OBJETOS EN MOVIMIENTO.....	10
2.1.1.	Pre-procesado.....	12
2.1.2.	Modelado de fondo.....	13
2.1.2.1.	Modelo de representación de fondo:.....	14
2.1.2.2.	Modelo de inicialización:.....	19
2.1.2.3.	Modelo de adaptación de fondo:.....	19
2.1.3.	Detección de frente.....	20
2.1.3.1.	Técnicas de detección de frente.....	20
2.1.3.2.	Selección del umbral.....	22
2.1.4.	Validación de datos o post-procesado.....	23
2.2.	ESTIMACIÓN DE MOVIMIENTO.....	23
2.2.1.	Principios generales de la estimación de movimiento.....	25
2.2.1.1.	La representación del movimiento.....	25
2.2.1.2.	Criterios de estimación.....	27
2.2.2.	Flujo óptico.....	33
2.2.2.1.	Ecuación del flujo óptico y ambigüedad en la estimación de movimiento.....	34
2.2.3.	Métodos basados en desplazamiento de bloques.....	37
2.2.3.1.	Región de exploración.....	38
2.2.3.2.	Elección del tamaño bloque.....	38
2.2.3.3.	Criterios de comparación.....	39
2.2.3.4.	Métodos de búsqueda.....	40
2.2.4.	Métodos basados en características.....	41
2.3.	MODELOS DE CLUSTERING.....	43
3.....		45
DISEÑO.....		45
3.1.	MODELO DE SEGMENTACIÓN IMPLEMENTADO.....	47
3.1.1.	Gaussiana Simple (SG).....	48
3.2.	MODELOS DE ESTIMACIÓN DE MOVIMIENTO IMPLEMENTADOS.....	51
3.2.1.	Método basado en el flujo óptico.....	52
3.2.1.1.	Normalización mediante la restricción de movimiento lento.....	53
3.2.1.2.	Restricción de región uniforme.....	54
3.2.2.	Métodos basados en desplazamiento de bloques implementados.....	57
3.2.2.1.	Algoritmo Block-Matching exhaustivo (EBMA).....	58
3.2.2.2.	Three Step Search (TSS).....	60
3.2.2.3.	New Three Step Search (NTSS).....	61
3.2.2.4.	Simple and Efficient TSS Search (SES).....	63

3.2.2.5.	Four Step Search (4SS)	65
3.2.2.6.	Diamond Search method.....	66
3.3.	MÉTODO DE CLUSTERING IMPLEMENTADO	68
3.3.1.	Algoritmo Mean Shift	68
4.	73
ALGORITMO PROPUESTO		73
4.1.	ETAPA DE SEGMENTACIÓN	74
4.2.	ETAPA DE ESTIMACIÓN DE MOVIMIENTO	77
4.3.	ETAPA DE AGRUPAMIENTO O CLUSTERING.....	81
4.4.	ETAPA DE SEGMENTACIÓN DE MOVIMIENTO	83
4.4.1.	Segmentación por movimiento inicial.....	86
4.4.2.	Subdivisión de regiones segmentadas	88
4.4.3.	Generación de la máscara de segmentación	93
5.	99
PRUEBAS Y RESULTADOS		99
5.1.	Descripción del dataset utilizado	100
5.2.	Diseño de pruebas experimentales.....	102
5.3.	Métricas utilizadas	106
5.4.	Selección de método óptimo de estimación de movimiento	108
5.4.1.	Tiempos de ejecución.....	109
5.4.2.	Calidad de la estimación de movimiento	112
5.4.3.	Conclusiones.....	113
5.5.	Selección de parámetros.....	114
5.5.1.	Parámetros para la segmentación	114
5.5.2.	Parámetros de la estimación de movimiento	115
5.5.3.	Parámetros óptimos para el clustering	115
5.5.4.	Parámetros óptimos para la segmentación de movimiento.....	116
5.6.	Análisis comparativo	117
5.6.1.	Calidad de la segmentación	117
5.6.1.1.	Resultados por secuencias	119
5.6.1.2.	Resultados globales.....	134
5.6.1.3.	Resultados: Conclusiones.....	140
6.	141
CONCLUSIONES Y TRABAJO FUTURO		141
6.1.	Trabajo futuro	143
BIBLIOGRAFÍA Y REFERENCIAS:		145
ANEXO 1		148
ANEXO 2		151

ÍNDICE DE FIGURAS:

Figura 1: Diagrama de las etapas de la segmentación	11
Figura 2: Máscara binaria tras el proceso de umbralización	22
Figura 3: Diferentes representaciones de movimiento: (a) Global.	27
Figura 4: Ejemplo de las esferas. (a) Esfera girando con foco de luz estático	33
Figura 5: Problema de la apertura. Solo se puede determinar con absoluta certeza el vector b	35
Figura 6: Moviendo la ventana de observación se pueden encontrar regiones exactamente iguales, que nos confunden a la hora de estimar el movimiento.	36
Figura 7: Distribución Gaussiana del píxel (188, 321) de una secuencia ejemplo en el fotograma 184.	48
Figura 8: Distribución Gaussiana de un píxel	49
Figura 9: Pirámide de imágenes. Cada nivel corresponde a una resolución diferente. ..	55
Figura 10: Sistema de búsqueda con bloque de tamaño MxN	59
Figura 11: Three Step Search.	61
Figura 12: Búsqueda NTSS. Puntos grandes pertenecen a primera iteración.	62
Figura 13: Patrones de búsqueda correspondientes a cada cuadrante:	64
Figura 14: Algoritmo SES. Vector resultado (3,-7).	64
Figura 15: Patrones de búsqueda en 4SS. (a) Primer paso.	66
Figura 16: Procedimiento de búsqueda 4SS. Punto óptimo (3,7).	66
Figura 17: Método Diamond Search. Este ejemplo muestra un punto óptico en (-4,-2) alcanzado en la quinta iteración.	67
Figura 18: Fase inicial del algoritmo	70
Figura 19: Segunda iteración. Se busca el punto óptimo.	70
Figura 20: Proceso finalizado con dos regiones encontradas	71
Figura 21: Diagrama de las etapas de la integración.	74
Figura 22: Ejemplo de máscara resultado de la segmentación basada en SG	76
Figura 23: Ejemplo de máscara Ground Truth.	77
Figura 24: Estimación de movimiento para los fotogramas 199 y 200 de la secuencia BAILE	79
Figura 25: Región de la estimación de movimiento	80
Figura 26: Máscara de segmentación mediante SG con valor $k = 3$	82
Figura 27: Resultado de Mean Shift sobre los vectores de movimiento de la imagen <i>enmascarada</i> con la máscara de la figura anterior.	82
Figura 28: Diagrama del proceso de clustering.	83
Figura 29: Segmentación mediante SG con $k = 2$	84
Figura 30: Segmentación mediante SG con $k = 3$	85
Figura 31: Segmentación mediante SG con $k = 4$	85
Figura 32: Clustering para $k = 2$	86
Figura 33: Clustering para $k = 3$	87
Figura 34: Clustering para $k = 4$	87
Figura 35: Esquema del primer proceso realizado para cada región.	88
Figura 36: Representación de la región 30	90
Figura 37: Representación de las subregiones de la región 30.	91
Figura 38: Representación de las subregiones de la región 30.	92
Figura 40: Representación de las subregiones de las subregiones obtenidas.	92
Figura 43: Diagrama de procesos realizados en MakeMask	94
Figura 44: Región 4 identificada como equivalente en la representación con $K=3$	96

Figura 45: Región 12 identificada como equivalente en la representación con $K=3$	96
Figura 46: : Región 14 identificada como equivalente en la representación con $K=3$...	97
Figura 47: Resultado de segmentación de movimiento iteración 6.....	98
Figura 48: Resultado de segmentación de movimiento iteración 13.....	98
Figura 49: Diagrama de barras de los parámetros estudiados en la secuencia Baile Unimodal	120
Figura 50: A la izquierda se muestra la máscara obtenida mediante el método basado en la <i>Gaussiana Simple</i> y a la derecha la obtenida con el método propuesto.	121
Figura 51: Diagrama de barras de los parámetros estudiados en la secuencia Baile Multimodal	123
Figura 52: A la izquierda se muestra la máscara obtenida mediante el método basado en la <i>Gaussiana Simple</i> y a la derecha la obtenida con el método propuesto.	124
Figura 53: Diagrama de barras de los parámetros estudiados en la secuencia Ingravidez.	129
Figura 54: A la izquierda se muestra la máscara obtenida mediante el método basado en la <i>Gaussiana Simple</i> y a la derecha la obtenida con el método propuesto.	130
Figura 55: Diagrama de barras de los parámetros estudiados en la secuencia Silla.....	132
Figura 56: A la izquierda se muestra la máscara obtenida mediante el método basado en la <i>Gaussiana Simple</i> y a la derecha la obtenida con el método propuesto.	133
Figura 57: Diagrama de barras de la precisión en cero por cada secuencia procesada para los dos métodos.	134
Figura 58: Diagrama de barras de la precisión en uno por cada secuencia procesada para los dos métodos	135
Figura 59: Diagrama de barras del <i>recall</i> en ceros por cada secuencia procesada para los dos métodos.	136
Figura 60: Diagrama de barras del <i>recall</i> en unos por cada secuencia procesada para los dos métodos.	137
Figura 61: Diagrama de barras del <i>Fscore</i> en ceros por cada secuencia procesada para los dos métodos.	138
Figura 62: Diagrama de barras del <i>Fscore</i> en unos por cada secuencia procesada para los dos métodos.	139
Figura 63: Diagrama de barras de la suma en unos por cada secuencia procesada para los dos métodos.	140

ÍNDICE DE TABLAS:

Tabla 1: Secuencias seleccionadas para el estudio	101
Tabla 2: Parámetros de prueba para la segmentación.....	103
Tabla 3: Parámetros de prueba para la estimación basada en Optical Flow.....	104
Tabla 4: Parámetros de prueba para la estimación basada en desplazamiento de bloques	104
Tabla 5: Relación entre parámetros de Mean Shift	105
Tabla 6: Rango de parámetros de Mean Shift para la agrupación de los vectores de movimiento.....	105
Tabla 7: Rango de parámetros de prueba para la segmentación de movimiento.	106
Tabla 8: Parámetros estadísticos de la evaluación.....	106
Tabla 9: Parámetros óptimos para la estimación de movimiento basada en BM	110
Tabla 10: Tiempo de procesamiento de los algoritmos de estimación de movimiento	111
Tabla 11: Número de puntos analizados para cada bloque	112
Tabla 12: Parámetros óptimos para las segmentación.....	115
Tabla 13: Relación entre parámetros de Mean Shift	116
Tabla 14: Parámetros óptimos para los procesos Mean Shift.....	116
Tabla 15: Parámetros óptimos para el proceso de segmentación de movimiento.	117
Tabla 16: Comparativa entre método con mejora y sin mejora.....	118
Tabla 17: Resultados para la secuencia Baile Unimodal.....	119
Tabla 18: Resultados para la secuencia Baile Multimodal.....	122
Tabla 19: Resultados para la secuencia Tennis	125
Tabla 20: Diagrama de barras de los parámetros estudiados en la secuencia Tennis. ...	126
Tabla 21: A la izquierda se muestra la máscara obtenida mediante el método basado en la Gaussiana Simple y a la derecha la obtenida con el método propuesto.	127
Tabla 22: Resultados para la secuencia Ingravidez	128
Tabla 23: Resultados para la secuencia Silla.....	131

1

INTRODUCCIÓN

1.1. MOTIVACIÓN

En la actualidad está cada vez más presente la necesidad de sistemas inteligentes para el análisis de secuencias de video, principalmente en los sistemas de video vigilancia. Ante la cantidad de información visual que se almacena por dichos sistemas, nace la necesidad de crear herramientas que nos sirvan para analizar automáticamente la información presente en los videos captados.

Una de las técnicas básicas en el análisis de secuencias de vídeo es la de diferenciar o separar del fondo de la escena (*background*) los objetos ajenos a él (*foreground*). Esta operación suele conocerse como segmentación. En el caso de una escena grabada por una cámara fija, las técnicas de segmentación más eficaces son las basadas en el modelado del fondo. Sin embargo, la mayor parte de estas técnicas no saca partido de la coherencia que todo objeto en movimiento debe mantener entre cuadros consecutivos de una secuencia de vídeo.

Para el estudio del movimiento de objetos o personas presentes en una secuencia de vídeo existen técnicas matemáticas que nos permiten conocer el punto de inicio

y el punto final de los píxeles homólogos pertenecientes a una pareja de imágenes consecutivas de una secuencia. El método más preciso para la estimación de movimiento es el basado en el flujo óptico, en el cual mediante operaciones con los gradientes espacio-temporales de la intensidad de la imagen se consigue una estimación del movimiento de cada uno de los píxeles, sin embargo es el método que mayor carga computacional requiere. Es por ello que el método más utilizado y estudiado es el algoritmo de *Block-Matching* (1) (2), que realiza una comparación entre bloques de las imágenes consecutivas. Dentro de *Block-Matching* existen modelos precisos que recorren las imágenes de manera exhaustiva (*Exhaustive Block-Matching Algorithm*), soportando por tanto una carga computacional muy alta, y métodos que realizan estudios menos precisos, a cambio de una mayor rapidez y simplicidad (*Fast Block-Matching Algorithm*). Existen otros métodos para la estimación de movimiento en secuencias de imágenes que en vez de realizar la comparación por bloques la realizan por regiones (*Region-Based Motion Estimation*) (3), mediante un estudio global de la imagen (*Global Motion Estimation*), o mediante un estudio de las imágenes una vez han sido divididas por mallas deformables (*Mesh-Based Motion Estimation*) (4). Por último, existen técnicas de estimación de movimiento basados en características. Éstos, identifican los puntos singulares de la imagen, tales como esquinas o líneas, y realizan una búsqueda para encontrarlos en la imagen sucesiva. De esta manera, teniendo un amplio conjunto de puntos, se puede obtener el movimiento que sufren los objetos de la imagen.

Una vez extraídos los vectores de movimiento de los distintos objetos segmentados se tiene el punto de partida para la integración de información de movimiento basado en el modelado de fondo. Para realizar dicha integración se debe, en primer lugar, evaluar si existe o no coherencia en el movimiento de los objetos segmentados. Para ello se acude al ajuste de conjuntos de vectores de movimiento a transformaciones geométricas que describan un movimiento similar. Los resultados de esta etapa permitirán modificar los resultados de la segmentación basada en modelado de fondo y, en su caso, modificar los parámetros que la rigen.

1.2. OBJETIVOS

La primera etapa del trabajo se centrará en el estudio de los métodos de segmentación existentes en el estado del arte. La segmentación es el proceso mediante el cual se discriminan los objetos en movimiento del frente (*foreground*) de una imagen, respecto a los del fondo (*background*). Se realizará una clasificación basada en tres modelos de representación del fondo. Existen métodos básicos (mediante diferencias, sumas, medias o medianas); modelos paramétricos, caracterizados por un conjunto de parámetros que describen la distribución de cada píxel; y los modelos no paramétricos, que están descritos por medio de funciones de probabilidad o otras funciones matemáticas. Se analizarán detalladamente cada uno de los métodos y se realizará una comparativa de los mismos.

Posteriormente se realizará un análisis exhaustivo a las técnicas de estimación de movimiento encontradas en el estado del arte por medio de los algoritmos que las definen, mediante los que se realizarán diferentes pruebas para escoger el método que mejores resultados ofrezca. Posteriormente se realizará un estudio sobre la coherencia del movimiento de los objetos o regiones extraídas a partir de la estimación del movimiento. Dicho estudio se basa en el análisis de los vectores de movimiento, y mediante algoritmos de parametrización se obtendrán tanto los valores correspondientes a los parámetros del movimiento, como una agrupación por regiones de los vectores que presenten un movimiento coherente entre sí. Para ello, se usarán métodos de *clustering*, que serán analizados a fondo más adelante. Una vez se obtenga una arquitectura capaz de arrojar resultados fiables, se realizarán pruebas en primera instancia con segmentación a través de sistemas *ground truth*, mediante los cuales se observarán los resultados ideales, necesarios para entender en profundidad la situación, y mediante los cuales se realizará la comparación con los sistemas reales, basados en los diferentes métodos de segmentación que se utilizarán.

Resumiendo, se podría definir la realización del proyecto en cuatro bloques, cada uno independiente entre sí, pero cuyo entrelazado produce un aumento de la calidad y la precisión de los resultados finales, complementándose hasta alcanzar un único modelo de segmentación de objetos en movimiento. Los dos primeros bloques quedarían definidos de la siguiente manera:

1. El método de **segmentación** permite agrupar los píxeles en regiones. La técnica empleado debe garantizar que cada región corresponda, si no a un objeto completo, sí al menos a una parte de la imagen perteneciente a un solo objeto. De esta manera, las regiones irán siendo agrupadas en la segmentación de movimiento para formar objetos.
2. El método de **estimación de movimiento** calculará el movimiento que cada punto o región ha experimentado en el plano de la imagen durante el tiempo transcurrido entre dos imágenes consecutivas de una secuencia. Con los parámetros estimados, la técnica de segmentación de movimiento (clustering) unirá aquellas regiones que se desplazan con un movimiento coherente, formando a su vez objetos o partes independientes de los mismos.
3. El tercer bloque consiste en un análisis de los vectores de movimiento. El análisis consiste en un agrupamiento, o **clustering**, de los vectores que presentan un movimiento coherente entre sí, con el fin de agruparlos por regiones de movimiento.
4. El cuarto y último bloque sería una integración de ambos métodos, con el fin de encontrar un método, basado en la recursividad, con el que se consiga mejorar los métodos iniciales de segmentación, ayudándonos por los resultados que ofrece la estimación de movimiento. Mediante una arquitectura realizada con éste fin, se analizará cómo se pueden mejorar las técnicas de segmentación propuestas, y cómo influyen en los resultados finales las variables que las definen.

Para obtener los objetivos requeridos, se organizará el trabajo de la siguiente manera, con el fin de esquematizar lo máximo posible la investigación para conseguir resultados independientes y ser capaces de enfocar los diversos estudios de la manera más precisa posible.

1. Estudio del estado del arte actual:
 - a. Estudio en profundidad de las técnicas de segmentación existentes en el estado de arte actual. Se estudiarán los métodos paramétricos y los no paramétricos, haciendo especial hincapié en los últimos.
 - b. Inicialmente, se realizará un estudio pormenorizado de las diferentes técnicas de estimación de movimiento. Se partirá de técnicas basadas en el *flujo óptico*, para acabar con los diversos métodos basados en el algoritmo *Block-Matching*.
2. Selección e implementación de los algoritmos más relevantes de estimación de movimiento y de segmentación: Una vez concluida la fase de recopilación de información y estudio, se pasará a la implementación de las técnicas estudiadas más precisas.
3. Desarrollo e implementación de un sistema de segmentación: Se desarrollará un sistema de evaluación de estimación de segmentación de objetos basada en los diferentes métodos de segmentación estudiados. Una vez obtenido este sistema se puede pasar al estudio real de los resultados obtenidos.
4. Desarrollo e implementación de un sistema de estimación de movimiento: Se desarrollará un sistema de evaluación de estimación de movimiento basada en los diferentes métodos estudiados. Una vez obtenido este sistema se puede pasar al estudio real de los resultados obtenidos.
5. Desarrollo e implementación de un sistema completo de segmentación de movimiento: Se desarrollará un sistema de evaluación global que aúne los

resultados obtenidos de los bloques anteriores por medio del estudio de los vectores de movimiento obtenidos.

6. Análisis de resultados y conclusiones: se extraerán una serie de conclusiones de cada algoritmo en función de la influencia de los diversos factores de las secuencias en los resultados y se ofrecerá una comparativa de los algoritmos

1.3. ORGANIZACIÓN DE LA MEMORIA

La memoria consta de los siguientes capítulos:

- **Capítulo 1:** Introducción, objetivos y motivación del proyecto.
- **Capítulo 2:** Antecedentes. Explicación de los métodos a utilizar, haciendo una descripción completa de los módulos y bloques en los que se dividirá el proyecto. Introducción a la segmentación: Modelos básicos. Modelos paramétricos. Modelos no paramétricos. Introducción a la estimación de movimiento. Introducción a los métodos de clustering
- **Capítulo 3:** Definición de los métodos de segmentación, de estimación de movimiento, y de clustering utilizados a lo largo de la realización del proyecto.
- **Capítulo 4:** Estudio de la arquitectura realizada. Estudio por bloques.
- **Capítulo 5:** Pruebas, resultados y análisis comparativo: descripción del conjunto de datos de prueba utilizado, pruebas y resultados.
- **Capítulo 6:** Conclusiones y trabajo futuro.

2

ANTECEDENTES

La percepción del movimiento es una función primordial en el sistema visual humano y animal, que juega un papel muy importante en tareas como percepción de profundidad o la detección y segmentación de objetos.

“Motion is powerful cue for image and scene segmentation in the human visual system as evidenced by the ease with which we see otherwise perfectly camouflaged creatures as soon as they move, and by the strong cohesion perceived when even disparate parts of the image move in a way that could be interpreted in terms of a rigid motion in the scene...” (5)

En los últimos tiempos, se ha prestado una especial atención al estudio de la percepción del movimiento por parte de los humanos (6). Con el fin de descubrir cómo funciona nuestro sistema visual, profesionales del mundo de la psicología, fisiología o visión por ordenador han centrado sus esfuerzos en entender como percibimos el movimiento e intentar formular teorías computacionales que recojan las explicaciones y las puedan aplicar en autómatas.

El análisis de movimiento consiste en extraer conocimiento a partir de una sucesión de imágenes. Para ello, se comparan las imágenes de la secuencia,

analizando las posibles similitudes o diferencias de las mismas. A partir de este análisis, existen dos tipos de problemas que a resolver. (7) (8).

1. *Detección de movimiento.* Consiste en detectar si se ha producido algún tipo de movimiento en la secuencia de imágenes.
2. *Detección y localización de objetos en movimiento.* Este problema es más complejo que el anterior, debido a que no solo consiste en saber si algo se ha movido, sino en saber la forma que tiene en el plano de la imagen. Se puede además requerir estimar la trayectoria de los objetos móviles o de las partes de los mismos que mantienen movimientos independientes, para obtener su posición futura. Para ello, se deben emplear técnicas de estimación de movimiento, segmentación, y en el caso más complejo, técnicas de seguimiento. Se entiende por *estimación de movimiento* en el plano de la imagen, el cálculo de los vectores de desplazamiento de los píxeles de la imagen. La estimación se puede realizar siguiendo un método exhaustivo, en el cual se recorren todos los píxeles de la imagen o dispersa, cuando sólo se realice la búsqueda en puntos concretos de la misma. Se debe tener en cuenta que el movimiento 2D, en el plano de la imagen, es una proyección del movimiento 3D experimentado por los objetos de la escena. La información obtenida al estimar el movimiento puede ser usada para determinar el número de objetos de la imagen, el número de regiones de cada objeto con movimientos independientes así como la velocidad de todos ellos. Para ello, los puntos que se mueven con parámetros de movimiento coherentes son agrupados en un proceso denominado *segmentación de movimiento*.

Las técnicas de estimación de movimiento han sido desarrolladas en infinidad de aplicaciones tales como la codificación de video en estándares de compresión; la visión artificial de robots; estudios meteorológicos basados en el seguimiento de las nubes; aplicaciones militares tales como búsqueda y reconocimiento de objetivos; biomedicina, en especial en el análisis del movimiento cardíaco y del movimiento de los órganos locomotores en la realización de ejercicio físico, con vistas a la rehabilitación de disfunciones congénitas; y en interfaces de realidad

virtual, mediante la interpretación de gestos y otras expresiones para interactuar con la interfaz.

El estudio del movimiento parte de los resultados ofrecidos por el proceso de segmentación, que tiene como objetivo la discriminación de los objetos en movimiento pertenecientes a una imagen (denominado frente o *foreground*), de aquellos objetos que permanecen estáticos (fondo o *background*). Al realizar previamente el proceso de segmentación, se evita realizar una estimación de movimiento en aquellas zonas de la imagen que no se identifican como *foreground*, evitando un exceso de carga computacional. Existen multitud de técnicas de segmentación de secuencias de video dependiendo de la naturaleza de las secuencias y los objetivos requeridos. Sin embargo la segmentación empleada en nuestro estudio está basada en el modelado y substracción de fondo, que consiste en un conjunto de algoritmos que mantienen un modelo matemático de apariencia del fondo a través del cual se pueden extraer los objetos en movimiento y el fondo de cada imagen de la secuencia.

La segmentación realiza la separación entre los puntos de la imagen estáticos y los dinámicos, pero hay una serie de situaciones ante las cuales los procesos de segmentación se ven en serios problemas para discernir con robustez si un determinado punto de la imagen permanece estático o no. Dichas situaciones pueden ser debidas a cambios de iluminación en la escena a analizar, sombras y reflejos en la escena, ruido introducido en las imágenes, o camuflaje, que consiste en una oclusión por parte objetos que poseen el mismo color que el objeto ocluido. Toyama (9) , Elgammal (10) y Harville (11).

Para todas las situaciones mencionadas se deben utilizar métodos lo suficientemente robustos como para obtener resultados coherentes. Para ello, se propondrá un proceso conjunto que consiste en el estudio del movimiento en dos tipos de escenarios. El primero mediante segmentación ideal, conseguida mediante un *ground truth* (12), en la cual ofrece una máscara ideal de los objetos que han experimentando algún movimiento entre dos o más imágenes consecutivas. Este

modelo servirá de referencia para todas las comparaciones que se realicen posteriormente, siendo el modelo que arroja unos resultados ideales. El otro escenario consistirá en un análisis de movimiento de las imágenes segmentadas mediante métodos de segmentación que no obtienen máscaras ideales. Los métodos se van a utilizar están basados en modelos básicos, paramétricos y no paramétricos, aunque se hará especial hincapié en los últimos, principalmente en la Gaussiana Simple (SG).

2.1. SEGMENTACIÓN DE OBJETOS EN MOVIMIENTO

A continuación se explicarán los conceptos básicos e imprescindibles para comprensión de la segmentación de objetos en movimiento. Estos conceptos son fundamentales para el posterior entendimiento de la arquitectura realizada en el proyecto, así como para tener una visión general del tema tratado.

El proceso inicial del proyecto consiste en la discriminación automática del frente de la imagen respecto al fondo. Como se ha explicado anteriormente hay diversas técnicas para la segmentación de objetos, dependiendo de los modelos de representación del fondo que se utilicen. Estos modelos hacen referencia al algoritmo matemático empleado en la representación del fondo de la escena, la zona de la imagen que permanece *estática*.

Se debe tener en cuenta que existen una serie de situaciones en las cuales los métodos de segmentación pueden tener problemas a la hora de discernir si un determinado píxel experimenta cambios en escenas sucesivas o no. (9) (10) (11). Se pueden tener cambios de iluminación en la escena a analizar, siendo dichos cambios provocados por fuentes naturales o artificiales, que pueden ser malinterpretados como parte de la imagen en movimiento, y catalogarse como frente. Las sombras y reflejos provocados por los objetos presentes en la escena también provocan errores en la detección, debido principalmente a que las sombras no son consideradas ni fondo ni frente, y deben eliminarse. Por otro lado se debe discernir qué sombras sí son inherentes al fondo para no eliminarlas y no

influir en la detección posterior. El ruido proveniente de la captación de las imágenes por las cámaras de video, también debe ser considerado para evitar fallos en la detección. Otro efecto interesante que resulta una fuente de problemas es el fenómeno del camuflaje, que aparece cuando los objetos del primer plano tienen el mismo color y textura que el fondo de la imagen, confundiéndose con el fondo y resultando difícil su detección.

Para realizar un estudio de los métodos de segmentación existentes en el estado del arte se tomará como punto de partida el Proyecto Fin de Carrera realizado por Sonsoles Herrero Martín, 'Análisis comparativo de técnicas de segmentación de secuencias de video basadas en el modelado del fondo' (13). Mediante el esquema realizado en dicho proyecto, se realizará una descripción básica de los métodos de segmentación, así como de los diferentes procesos utilizados para su definición. Se puede dividir la segmentación en cuatro etapas, definidas inicialmente por Cheung y Kamath (14) y mostradas en la Figura 1:

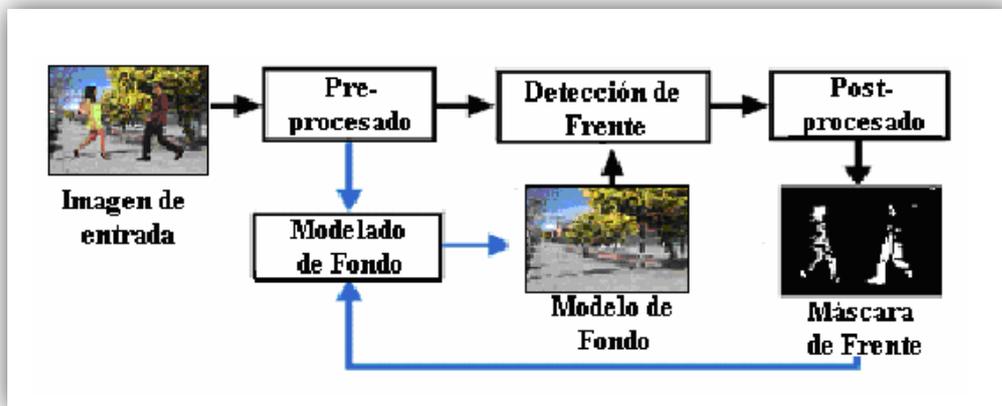


Figura 1: Diagrama de las etapas de la segmentación

2.1.1. Pre-procesado

Su función principal consiste en filtrar los cambios significativos que ocurren en la escena y mejorar las propiedades de las imágenes de la secuencia a analizar. Estos procesos contribuyen a la simplificación de la etapa posterior consistente en la detección de los objetos. En esta etapa se suelen utilizar una serie de ajustes geométricos o ajustes de intensidad (15), aunque también existen otro tipo de técnicas que utilizan la derivación como medio para la obtención de bordes y contornos, que resultan fundamentales en el reconocimiento de objetos.

A continuación se describen las técnicas de pre-procesado más utilizadas en el estado del arte.

- ***Ajustes geométricos:*** Las técnicas de ajustes geométricos permiten la regeneración de los píxeles que han sido afectados por cambios en la intensidad o movimientos en la cámara, mediante operaciones con los valores de píxeles adyacentes. Los principales ajustes geométricos son la dilatación y erosión, la rotación, y las transformaciones afines (15) (16).
- ***Ajustes de intensidad o radiométricos:*** Métodos muy útiles para compensar las variaciones que se pueden producir en los píxeles debido a cambios en la intensidad o posición de las fuentes de iluminación. El ajuste puede consistir en una normalización de los píxeles de la imagen para que tengan la misma media y varianza que los píxeles del modelo de fondo.
- ***Derivada de la imagen:*** Estas técnicas de pre-procesado calculan la derivada de la imagen (diferencia de frentes) para general una representación de la escena de fondo invariante a los cambios de iluminación o para detectar objetos que presentan movimientos a alta velocidad.

2.1.2. Modelado de fondo

Es la etapa más importante en la segmentación de objetos en movimiento. Su función principal es la inicialización, actualización y representación de un modelo de fondo de la secuencia analizada. El objetivo, por lo tanto no es encontrar información relativa a los objetos en movimiento de la imagen, ya que se realizará en etapas posteriores. En los casos más sencillos, entornos controlados en los que la imagen de fondo es conocida y estática, denominados *fondos unimodales*, se puede realizar una aproximación consistente en suponer que los píxeles del fondo no cambian de valor a lo largo de la secuencia y sólo los píxeles de los objetos en movimiento cambian de valor. En cambio, para los entornos menos controlados incluso los píxeles del fondo de la imagen cambian de valor debido a cambios en la iluminación (escenas a la luz del día, donde se producen variaciones naturales) o al movimiento de objetos que formen parte del fondo (escenas en las cuales aparezcan árboles). Este tipo de fondo se conocen como *fondos multimodales*.

Para realizar un correcto modelado de fondo, se deben conocer una serie de características de la secuencia, como pueden ser si se trata de un fondo unimodal o multimodal, a qué velocidad se están moviendo los objetos o con qué velocidad se producen cambios de iluminación en la imagen. Todo ello se debe tener en cuenta a la hora de escoger cada uno de los métodos dependiendo de la situación a estudiar.

Según Cristani (17), la etapa de modelado de fondo posee tres características fundamentales:

1. **Modelo de representación:** describe el modelo matemático utilizado para la representación del fondo.
2. **Modelo de inicialización:** se refiere a la forma con la que se obtiene el primer fondo o estado del modelo del algoritmo.

3. **Modelo de adaptación:** mecanismo para adaptar el modelo a cambios en el fondo de la escena (por ejemplo, cambios de iluminación).

A continuación, se analizarán una serie de técnicas existentes en base a las tres características descritas.

2.1.2.1. **Modelo de representación de fondo:**

Este modelo describe el modelo matemático utilizado para la representación del fondo. La definición de los distintos modelos de representación será fundamental en el resto del proyecto, y se hará especial hincapié en su definición en el capítulo referente a la arquitectura realizada. Para realizar una breve introducción de los métodos, se realizará una explicación pormenorizada de qué consiste cada una de las tres clases de modelos estudiados:

- **Métodos básicos:**

Métodos matemáticos sencillos para generar el fondo de la imagen, separándolo del frente, tales como diferencias entre imágenes, promedios, máximos y mínimos, etc..., que permiten modelar de manera simple los píxeles de la imagen de fondo. A continuación se describen alguno de estos métodos:

- ***Diferencia de imágenes***

El método de diferencia de imágenes (18), también denominado diferencia temporal, es posiblemente la forma más sencilla de sustracción de fondo.

Utiliza como modelo de fondo (B_t) para la imagen de la secuencia en el instante (I_t) la imagen anterior, es decir, la imagen en el instante ($t - 1$).

Esta técnica posee una baja carga computacional, y el frente tiene buena adaptación a los cambios rápidos. Por otra parte es sensible al ruido y a las variaciones de iluminación, así como resulta ineficaz en la detección del interior de

los objetos en movimiento, puesto a que entre imágenes consecutivas, esos píxeles se mantienen invariantes.

- ***Filtro promedio temporal***

En el filtro de promedio el modelo de fondo se calcula como una imagen estática hasta que se producen movimientos, en cuyo caso, el B_t corresponde al promedio de un conjunto de imágenes consecutivas en el tiempo.

$$|I_t - B_t| > \tau \Rightarrow B_t = \frac{1}{N} \sum_{n=t}^{n=t-N} I_n$$
$$|I_t - B_t| \leq \tau \Rightarrow B_t = I_t$$

Este modelo no es robusto con muchas secuencias en movimiento, sobre todo si se mueven lentamente. Además no puede manejar fondos multimodales, adapta lentamente las variaciones de fondo y posee un único umbral para toda la imagen.

- ***Filtro de mediana aproximado***

En este método se almacena en un buffer los valores de los píxeles en un conjunto de imágenes. El fondo se define en cada píxel como la mediana de los valores almacenados en el buffer. Es un método complejo cuyos resultados se asemejan a los de los métodos complejos.

$$I_t > B_t \Rightarrow B_t = B_{t-1} + 1$$
$$I_t < B_t \Rightarrow B_t = B_{t-1} - 1$$
$$I_t = B_t \Rightarrow B_t = B_{t-1}$$

La estimación converge a un valor tal que la mitad de los píxeles de entrada son mayores a él y la otra mitad son inferiores. Es un estimador de la mediana.

- **Métodos paramétricos:**

Los algoritmos basados en modelos paramétricos definen modelos de fondo más complejos, que permiten cierta tolerancia al ruido y a pequeñas fluctuaciones tales como hojas en movimiento, parpadeo de luces, vibraciones de las cámaras, etc. Describen la imagen de fondo en base a parámetros de una distribución de probabilidad estándar (normalmente Gaussiana). Algunos ejemplos son los mostrados a continuación:

- ***Gaussiana Simple ('Simple Gaussian')***

Este método modela los pequeños cambios que ocurren en la imagen de fondo B_t representando cada píxel con una distribución unimodal Gaussiana definida por dos parámetros: media μ_t y la varianza σ_t^2 .

$$\mu_t(x, y) = \sum_{i=1}^{i=t} \frac{B_i(x, y)}{t}$$
$$\sigma_t^2(x, y) = \sum_{i=1}^{i=t} \frac{B_i^2(x, y)}{t} - \mu_t^2(x, y)$$

En cada instante t , se determina si un píxel pertenece al fondo $B_t(x, y)$ si el valor de dicho píxel en la imagen recae dentro de la Gaussiana definida para ese píxel, es decir, si la diferencia entre el valor del píxel y el de la media considerada para dicho píxel $\mu_t(x, y)$ es inferior a la varianza $\sigma_t^2(x, y)$.

El filtro de Kalman es una técnica recursiva de seguimiento lineal en la que los valores de los píxeles pueden modelarse por una distribución Gaussiana $N(\mu, \sigma_t^2)$. La media y la varianza del fondo se actualizan utilizando filtros adaptativos para acomodar los cambios de iluminación.

- **Mezcla de Gaussianas ('Mixture of Gaussian')**

Existen fondos multimodales, que contienen fondos no estáticos, en los cuales hay píxeles cuyos valores de intensidad varían entorno a un conjunto finito de valores característicos. Por ello, un píxel no puede modelarse por medio de un único valor (una media) y un conjunto en torno a éste (la varianza) utilizando una distribución Gaussiana. Este método propone una solución al problema (19) que consiste en modelar la intensidad de los píxeles con una mezcla de k distribuciones Gaussianas (siendo k un valor de 3 y 5) definidas por los siguientes parámetros: media $\mu_{k,t}$, varianza $\sigma_{k,t}^2$ y peso $w_{k,t}$.

$$\mu_{k,t}(x, y) = \sum_{i=1}^{i=t} \frac{B_i(x, y)}{t}$$
$$\sigma_{k,t}^2(x, y) = \sum_{i=1}^{i=t} \frac{B_i^2(x, y)}{t} - \mu_t^2(x, y)$$

Con este método, se comparan los píxeles de una imagen nueva con las k posibles distribuciones de fondo; si el píxel no difiere de alguna de las medias en más de la varianza asociada, se caracteriza el píxel como fondo y se actualizan los parámetros del modelo para ese píxel: la media, la varianza y un peso cuyo valor determina la cantidad de distribución que modela dicho píxel en cada instante. En el caso de que no se parezca a ninguna de las distribuciones asociadas a cada píxel, la de menor peso se sustituye por una nueva Gaussiana de media el valor del píxel de la imagen actual y varianza un valor muy pequeño.

- **Métodos no paramétricos:**

En este caso la representación del fondo se realiza por medio de funciones de densidad de probabilidad y otras funciones matemáticas más generales, pero sin asumir distribuciones concretas que podrían ser definidas por parámetros. Algunos modelos paramétricos son los siguientes:

- **Densidad de Núcleo ('Kernel Density Estimate')**

El método de representación del fondo de Densidad de Núcleo ('KDE'), enunciado por Elgammal (10) estima la función de densidad de probabilidad de cada píxel de la imagen de fondo en cada instante de tiempo, gracias a la información de la historia reciente de dicho píxel que ha sido almacenada en un buffer.

$$B_t(x, y) = \text{buffer}\{I_{t-L}(x, y), I_{t-L+1}(x, y), I_t(x, y)\}$$

En este caso, la pertenencia al fondo B_t se estima mediante el promedio ($\text{Pr}(x_t)$) de funciones de núcleo K (por ejemplo, de tipo Gaussiano) centradas en valores anteriores x_i y evaluados en el valor de píxel actual x_t . En otras palabras, se calcula la probabilidad del parecido entre el píxel actual y los valores que el píxel ha tomado en las L imágenes anteriores, y si esa probabilidad es suficientemente elevada se considera que el píxel es fondo.

El modelo KDE soporta parpadeo de fondo, ruido en la imagen y es capaz de adaptarse a los cambios rápidos y progresivos del fondo. No obstante, posee un requisito muy importante, que es la alta carga computacional.

- **Modelos ocultos de Markov ('HMM')**

Otro enfoque para modelar una amplia gama de variaciones en la intensidad sería utilizando los modelos ocultos de Markov (HMM), que representan las variaciones como un conjunto de estados discretos correspondientes a los distintos modos de iluminación que pueden presentarse en la escena.

El modelo de fondo B_t quedaría definido por un conjunto de estados b_j caracterizados por una función de densidad de probabilidad para cada píxel localizado en la posición (x,y) :

$$B_t(x, y) = b_j(x, y)$$

Este modelo requiere un tiempo de cálculo elevado ya que suponen la evaluación de un conjunto de estados y la tipología del sistema puede resultar muy compleja.

2.1.2.2. Modelo de inicialización:

La mayoría de los modelos de fondo se basan en la generación inicial de un fondo mediante una serie de parámetros obtenidos a través de un número determinado de imágenes de la secuencia en la que no estén presentes los objetos en movimiento (20). Esta simplificación es difícil o imposible de controlar en la mayoría de los casos, con lo que nos resulta imprescindible la utilización de un modelo que nos proporcione fondos previos robustos.

2.1.2.3. Modelo de adaptación de fondo:

Es necesaria la actualización del modelo de fondo a los cambios que puedan producirse en las secuencias analizadas, tales como cambios en la iluminación o nuevos objetos que aparecen en el escenario. Para adaptar el fondo, existen diversas técnicas de adaptación. Una manera sencilla sería una actualización periódica (cada instante) del modelo a través de la media o mediana de los valores de las imágenes anteriores. Técnicas más complejas mantienen un buffer para almacenar los valores de cada píxel a lo largo de un periodo determinado, actualizando el modelo en función de los valores del buffer.

Las técnicas de actualización del modelo de fondo pueden realizarse mediante dos mecanismos diferentes: mecanismo de adaptación selectiva o a ciegas.

- ***Mecanismo de adaptación selectiva:*** Este método selecciona y actualiza sólo los píxeles de la imagen clasificados como fondo, permitiendo incorporar los objetos que aparezca en la escena y que pasado un determinado tiempo corresponden al modelo de fondo.

- **Mecanismo de adaptación a ciegas:** Este método actualiza todos los píxeles del modelo de fondo, sin tener en cuenta ningún tipo de decisión previa. Por ello, estos métodos requieren técnicas de actualización de menor complejidad que los mecanismos selectivos, pero a cambio presentan como desventaja la actualización del fondo con los objetos en movimiento indeseados.

2.1.3. Detección de frente

Consiste en la detección del frente o primer plano, en la que se compara la entrada con el modelo de fondo y se identifican los posibles objetos en movimiento. La comparación puede realizarse por diferencia mientras que la clasificación se realiza mediante un umbral de decisión, que depende de la escena, el ruido de la cámara y las condiciones de iluminación.

Se pueden definir una serie de técnicas de detección de frente distintas, así como diversos métodos de selección de umbral. Se realizará una pequeña clasificación:

2.1.3.1. Técnicas de detección de frente

Son algoritmos sencillos que se aplican al modelo de fondo realizado en etapas anteriores. Realizan la comparación entre la imagen actual y el fondo, y detectan los objetos en movimiento. Existen dos tipos de técnicas, basadas en diferencia y basadas en estadísticas:

- **Basadas en diferencia:**

Es el método más básico, consisten en calcular la diferencia entre dos imágenes: los píxeles cuyos valores de diferencia son elevados corresponderán a objetos en movimiento. Esta diferencia puede calcularse por diversos métodos: *diferencia absoluta*, mediante el valor absoluto de la diferencia; *diferencia relativa*, en la que se aplica una diferencia en función de la distribución espacial de localización (21); *diferencia normalizada*, en el que se aplica el umbral sobre las estadísticas

realizadas; o la *diferencia basada en dos umbrales*, mediante la cual se aumenta el grado de precisión en la detección de frente.

- **Basadas en estadísticos:**

Estas técnicas únicamente pueden utilizarse en el caso de que se conozca la función de densidad de ruido de la cámara. Aquí, en vez de umbralizar la diferencia de la imagen, se compara el comportamiento estadístico de un conjunto de valores de un píxel perteneciente a una serie de imágenes de la secuencia con el mismo píxel de la imagen actual. Estos métodos dependen de la distribución con la que se modele el fondo, con lo que cada método utiliza una estrategia diferente para detectar los píxeles del fondo. Los algoritmos más populares son:

- ***Basados en una Gaussiana***, en los que la imagen actual se compara con el modelo de fondo de distribución Gaussiana (μ_i, σ_i^2) mediante la medición de la probabilidad en luminancia o color, y si la probabilidad es pequeña, el píxel se clasifica como frente, o de lo contrario, como fondo.
- ***Basado en Mezcla de Gaussianas (MoG)***: El valor de cada píxel se compara con las k distribuciones que modelan el píxel en una MoG (μ_i, σ_i^2) para decidir si el elemento pertenece al fondo (si recae dentro de la Gaussiana k-ésima) o al frente (si la diferencia supera la varianza permitida)
- ***Basado en la estimación de densidad de núcleo e Histogramas***: En este caso se generan estimaciones de probabilidad para cada píxel de la imagen en función de los píxeles almacenados de un número de imágenes anteriores a la procesada, de modo que, un píxel se considera primer plano si la probabilidad de pertenecer a esta estimación es inferior a un umbral.

2.1.3.2. Selección del umbral

Esta selección es una operación decisiva en la búsqueda de objetos de interés en la secuencia de análisis. Como resultado del cálculo de la imagen diferencia, los niveles de gris pertenecientes a los píxeles de frente tendrán valores superiores a los píxeles en ausencia de movimiento, por lo que el problema de localización de objeto en movimiento se resolverá buscando un umbral óptimo. La salida de una operación de umbralización será una máscara binaria, con valor 0 (equivalente al negro) si pertenece al fondo y 1 (blanco) si pertenece al frente, como se observa en la siguiente figura.



Figura 2: Máscara binaria tras el proceso de umbralización

La eficiencia del cálculo del frente depende de la selección del umbral, con lo que existen diversos métodos para establecer el umbral. Se puede realizar un establecimiento empírico, (22) (23) (24), en el que se fija un umbral para todos los píxeles, o a través de un cálculo adaptable, (25) (26) (27), cuyo umbral se adapta siguiendo diversos criterios, como pueden ser las características del modelo de fondo, o la información espacial.

2.1.4. Validación de datos o post-procesado

Existen aplicaciones en las que la detección requerida debe ser muy precisa para lograr objetivos claros, como puede ser el conteo de personas, la detección de objetos robados o abandonados, seguimiento de objetos en movimiento o análisis de comportamientos. En general, para obtener resultados fiables se debe realizar una etapa de post-procesado con el fin de reducir los problemas que pueden afectar a la detección y permitan simplificar la labor de los algoritmos posteriores. En la mayoría de los casos es necesaria eliminar el ruido presente en los resultados, corregir la falta de iluminación uniforme, y eliminación de sombras y reflejos de los objetos.

2.2. ESTIMACIÓN DE MOVIMIENTO

En este punto del capítulo se enunciarán los principios fundamentales de la estimación de movimiento, que serán el punto de partida para el análisis más profundo que se realizará en el capítulo de diseño, en el que se recogerán los métodos que se han llevado a la práctica en este proyecto. Es importante recoger en este punto del capítulo de antecedentes, todos los principios teóricos de la estimación de movimiento, con el propósito de ofrecer al lector una base de conocimiento suficiente para el entendimiento de los capítulos siguientes.

Para el estudio de secuencias de video debe saberse que el análisis de movimiento proporciona una información fundamental a un sistema basado en la visión. Con la ayuda de la información de movimiento es fácil interpretar una escena, permitiendo detectar e identificar las partes de los objetos que se hallan en movimiento, así como observar qué tipo de movimiento experimentan.

Todos los algoritmos de estimación de movimiento están basados en los cambios temporales en el color (intensidad) que se producen en imágenes consecutivas. De

hecho, los movimientos bidimensionales basados en la intensidad pueden no coincidir con los movimientos reales. Para ser precisos, la velocidad de los vectores bidimensionales observados o aparentes es denominada *flujo óptico*. Dicho flujo óptico no solamente es causado por el movimiento de los objetos, sino también por movimientos de la cámara o por cambios en la iluminación.

Existen una gran cantidad de métodos de análisis de movimiento, usualmente divididos en tres grandes grupos (28) (29) (30):

1. *Métodos basados en el gradiente*: Realiza una estimación de movimiento para todos y cada uno de los píxeles de la imagen, mediante el estudio de los gradientes espacio-temporales de la intensidad de la imagen a analizar. Para ello se escoge la denominada ecuación del flujo óptico para la obtención de la estimación de movimiento. Es el método más preciso para nuestro cometido, pero presenta un alto coste computacional.
2. *Métodos basados en el desplazamiento de bloques*: Realiza una estimación de movimiento mediante un algoritmo de búsqueda para obtener el bloque de la imagen sucesiva que tiene mayor semejanza con el bloque buscado de la imagen actual. Existen una serie de métodos, denominados *Block-Matching* mediante los cuales se realizan iteraciones que comparan el parecido de un bloque con los de su alrededor. Existen métodos exhaustivos, que comparan uno por uno todos los bloques, y métodos que simplifican el procedimiento de cálculo, realizando comparaciones más efectivas. Como se puede suponer, la precisión del método escogido, conlleva un aumento en el coste computacional del proceso, siendo interesante la comparación entre métodos, para su elección dependiendo de los requisitos del sistema.
3. *Métodos basados en características*: Realizan una búsqueda en la que se pretende relacionar puntos característicos de la imagen actual, tales como esquinas o bordes, con los mismos puntos de la imagen siguiente. Para ello hay que realizar un estudio de los puntos, y posteriormente una búsqueda de equivalencias.

2.2.1. Principios generales de la estimación de movimiento

A continuación se van a explicar las distintas representaciones con las que se puede mostrar el movimiento que existe en una secuencia de imágenes. Los distintos métodos aquí mencionados deben ser utilizados teniendo en cuenta tanto las características de las imágenes, como las del movimiento presente en las mismas. Tras esta explicación, se introducirán los principios matemáticos en los que se basa la estimación del movimiento, estudiando los distintos criterios mediante los que se puede encontrar el punto de la imagen actual correspondiente con el mismo punto de la imagen previa.

2.2.1.1. La representación del movimiento

Cualquier campo desplazamiento resultante del movimiento de la cámara o de los objetos puede ser descrito por una serie de parámetros. Un problema importante en la estimación de movimiento es cómo parametrizar dichos campos. Al realizar el estudio del movimiento de objetos puede suceder que en una imagen pueden existir una serie de objetos que mantienen movimientos independientes, con lo cual, para estos casos, un movimiento global resulta inadecuado.

El método más directo e intuitivo consiste en especificar un vector de movimiento para cada píxel, denominado *píxel-based representation*, o lo que se conoce por flujo óptico o estimación de movimiento basada en el gradiente. La representación es universalmente aplicable, pero presenta un coste computacional alto, al realizar una estimación de un gran número de incógnitas (dos veces el número de píxeles) y la solución puede resultar físicamente incorrecta a menos que se impongan las restricciones anteriormente explicadas, en la fase de estimación. Por otra parte, cuando únicamente es la cámara la que se mueve, la escena captada contiene un único objeto en movimiento con una superficie plana, se puede utilizar la *global motion representation* para caracterizar el movimiento de la escena. Pero en general, cuando las imágenes contienen múltiples objetos en movimiento, lo más

apropiado es dividir la imagen en diversas regiones para conseguir que cada región pueda ser estimada de modo independiente para conseguir modelos de parametrización precisos. Este método es conocido como *region-based motion representation*, (representación basada en regiones), el cual consiste en una división del mapa de segmentación y varios grupos de parámetros correspondientes a cada una de las regiones. La principal dificultad de este método es que se debe tener conocimiento previo de qué píxeles mantienen movimientos similares. Por lo tanto, la segmentación y la estimación de movimiento deben realizarse iterativamente, lo que significa un coste computacional en ocasiones excesivo en la práctica.

Un modo de reducir la complejidad asociada a la *region-based representation*, es por medio de la división del dominio de la imagen a estudio en pequeños bloques de tamaño fijo. Mientras que el tamaño de los bloques sea suficientemente pequeño, la variación del movimiento de cada bloque puede ser caracterizada de manera precisa mediante a un modelo simple, y la estimación de movimiento de cada bloque puede estimarse independientemente. Dicho método se denomina *Block-Matching* (representación basada en correspondencia de bloques). Éste proporciona un gran compromiso entre la precisión y la complejidad, y ha conseguido un gran éxito en sistemas prácticos de codificación de video. Un problema importante con este método *block-based* es que no impone ninguna restricción en la transición entre bloques adyacentes, con lo que el movimiento resultante es, en ocasiones, discontinuo a lo largo de las fronteras entre bloques, incluso cuando el movimiento real cambia suavemente de bloque en bloque. Un modo de resolver dicho problema es mediante la *mesh-based representation* (representación basada en mallas), en la cual la imagen en cuestión es dividida en elementos poligonales no solapados. El movimiento de la imagen queda de esta manera descrito por el movimiento de los nodos (esquinas de los polígonos) y el movimiento en el interior de las regiones se determina mediante la interpolación de los mismos. Esta representación arroja un movimiento completamente continuo en toda la imagen. Este método es más apropiado que el método del bloques en el interior de los objetos, que normalmente mantienen movimientos continuos, pero falla en el caso de límites y discontinuidades en los objetos.

La siguiente figura nos muestra los efectos de varias representaciones de movimiento:

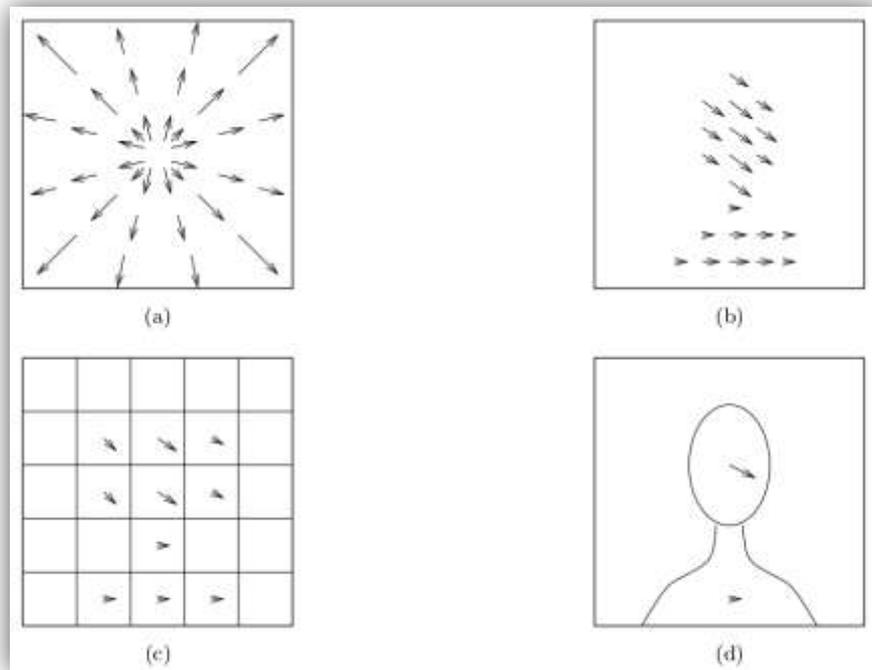


Figura 3: Diferentes representaciones de movimiento: (a) Global. (b) Basada en píxeles. (c) Basada en bloques. (d) basada en regiones

2.2.1.2. Criterios de estimación

Existen una serie de modelos matemáticos, que nos son de gran utilidad para la estimación del movimiento, es decir, para obtener valores reales del movimiento que experimente una región o punto de la imagen. Para ello, se tienen diferentes técnicas, que serán incluidas en los algoritmos de estimación, que nos informan de las diferencias de intensidad entre un punto determinado de la imagen previa y el mismo punto, u otros en el caso de la búsqueda por regiones. Debido a esta parametrización, se obtiene un valor de cuánto se diferencia un punto entre dos instantes de tiempo.

Para formular estos criterios de estimación, nos pueden ser útiles una serie de representaciones del movimiento y las relaciones entre movimiento e imágenes que se ha visto en el punto anterior. La dificultad para establecer un buen criterio reside principalmente en el hecho que el movimiento en imágenes puede no ser siempre observable y ese cambio en la intensidad de una secuencia puede ser el resultado de más de un movimiento. Otro problema es que la mayoría de los modelos analizados anteriormente están lejos del modelo ideal debido a que las diferencias de intensidad entre imágenes consecutivas pueden variar debido al ruido, a superficies no opacas, oclusiones o variaciones espacio-temporales de iluminación.

Criterio basado en la Diferencia de la Imagen Desplazada (DFD)

El criterio basado en la imagen desplazada es el criterio más popular debido a su simplicidad. Se basa en la suma de las diferencias entre los valores de luminancia de cada par de puntos correspondientes entre la imagen previa Ψ_1 y la actual Ψ_2 . Teniendo en cuenta que el punto \mathbf{x} en Ψ_1 se mueve a $\mathbf{w}(\mathbf{x}; \mathbf{a})$ en Ψ_2 , la función puede ser escrita como:

$$E_{DFD}(\bar{\mathbf{a}}) = \sum_{x \in \Lambda} |\psi_2(\bar{\mathbf{w}}(\bar{\mathbf{x}}; \bar{\mathbf{a}})) - \psi_1(\bar{\mathbf{x}})|^p, \quad (\text{Ec.1})$$

donde Λ es el conjunto de píxeles en Ψ_1 , y p es un número positivo. Cuando $p=1$, el error más alto es llamado **error medio absoluto (MAD)**, y cuando $p=2$, **error cuadrático medio (MSE)**. El error imagen $e(\bar{\mathbf{x}}; \bar{\mathbf{a}}) = |\psi_2(\bar{\mathbf{w}}(\bar{\mathbf{x}}; \bar{\mathbf{a}})) - \psi_1(\bar{\mathbf{x}})|$, es llamado normalmente **displaced frame difference (DFD) image**, y el sumatorio de las imágenes es llamado **DFD error**.

La condición necesaria para la minimizar E_{DFD} es que su gradiente tienda a cero. En el caso de $p=2$, el gradiente será:

$$\frac{\partial E_{DFD}}{\partial \bar{\mathbf{a}}} = 2 \sum_{x \in \Lambda} (\psi_2(\bar{\mathbf{w}}(\bar{\mathbf{x}}; \bar{\mathbf{a}})) - \psi_1(\bar{\mathbf{x}})) \frac{\partial \bar{\mathbf{d}}(\bar{\mathbf{x}})}{\partial \bar{\mathbf{a}}} \nabla \psi_2(\bar{\mathbf{w}}(\bar{\mathbf{x}}; \bar{\mathbf{a}})) \quad (\text{Ec.2})$$

Donde

$$\frac{\partial \vec{d}}{\partial \vec{a}} = \begin{bmatrix} \frac{\partial d_x}{\partial a_1} & \frac{\partial d_x}{\partial a_2} & \dots & \frac{\partial d_x}{\partial a_L} \\ \frac{\partial d_y}{\partial a_1} & \frac{\partial d_y}{\partial a_2} & \dots & \frac{\partial d_y}{\partial a_L} \end{bmatrix}^T \quad (\text{Ec.3})$$

Criterio basado en la ecuación de flujo óptico

En vez de minimizar el error DFD, otra aproximación es resolver el sistema de ecuaciones establecido basado en la restricción del flujo óptico.

Considerando una secuencia de video cuyo brillo queda representada por $\psi(x, y, t)$ (cabe destacar que se considera movimiento en base a la información intensidad de luz). Se supone que un punto de la imagen (x, y) perteneciente al instante t es movido a $(x + d_x, y + d_y)$ en el instante $t + d_t$. Mediante la restricción de intensidad constante, (31), las imágenes de los puntos del mismo objeto en instantes diferentes tienen la misma intensidad luminosa. Con lo cual:

$$\psi(x + d_x, y + d_y, t + d_t) = \psi(x, y, t) \quad (\text{Ec.4})$$

Mediante la expansión de Taylor, cuando $d_x, d_y, y d_t$ son valores pequeños, se tiene que:

$$\psi(x + d_x, y + d_y, t + d_t) = \psi(x, y, t) + \frac{\partial \psi}{\partial x} d_x + \frac{\partial \psi}{\partial y} d_y + \frac{\partial \psi}{\partial t} d_t \quad (\text{Ec.5})$$

Y combinando (Ec.4) y (Ec.5) se obtiene:

$$\frac{\partial \psi}{\partial x} d_x + \frac{\partial \psi}{\partial y} d_y + \frac{\partial \psi}{\partial t} d_t = 0. \quad (\text{Ec.6})$$

La ecuación (Ec.6) puede ser escrita en función del vector de movimiento (d_x, d_y) . Dividiendo ambos miembros por el diferencial temporal d_t , se obtiene la ecuación en términos de velocidad:

$$\frac{\partial \psi}{\partial x} v_x + \frac{\partial \psi}{\partial y} v_y + \frac{\partial \psi}{\partial t} = 0 \quad \text{ó} \quad \nabla \psi^T \vec{v} + \frac{\partial \psi}{\partial t} = 0 \quad (\text{Ec.7})$$

donde (v_x, v_y) representan los *vectores de velocidad* (también denominados *vectores de flujo óptico*), y $\nabla \psi = \left[\frac{\partial \psi}{\partial x}, \frac{\partial \psi}{\partial y} \right]^T$ es el gradiente espacial de $\psi(x, y, t)$.

Esta ecuación, comúnmente llamada **Ecuación del Flujo Óptico**, relaciona el campo velocidad (v_x, v_y) con los gradientes o variaciones espacio-temporales de la intensidad de la imagen $\partial \psi / \partial x, \partial \psi / \partial y, \partial \psi / \partial t$.

El vector de velocidad \mathbf{v} en el punto \mathbf{x} puede ser descompuesto en dos componentes ortogonales:

$$\vec{v} = v_n \vec{e}_n + v_t \vec{e}_t, \quad (\text{Ec.8})$$

donde \vec{e}_n es la dirección del vector gradiente de la imagen $\nabla \psi$, denominado dirección normal y \vec{e}_t , que es ortogonal a \vec{e}_n , es denominado dirección tangente. La ecuación del flujo óptico (Ec.4) puede ser escrita de la siguiente manera:

$$v_n \|\nabla \psi\| + \frac{\partial \psi}{\partial t} = 0, \quad (\text{Ec.9})$$

donde $\|\nabla \psi\|$ es la magnitud del vector gradiente $\nabla \psi$.

Llegado a este punto, se debe retomar la ecuación Ec.6. Dado $\psi_1(x, t) = \psi(x, y, t)$ y $\psi_2(x, y) = \psi(x, y, t + d_t)$. Si d_t es pequeña, se asume que $(\partial \psi / \partial t) dt = \psi_2(\vec{x}) - \psi_1(\vec{x})$. Con lo que la ecuación (Ec.6) puede escribirse como:

$$\frac{\partial \psi_1}{\partial x} d_x + \frac{\partial \psi_1}{\partial y} d_y + (\psi_2 - \psi_1) = 0 \quad \text{ó} \quad \nabla \psi_1^T \vec{d} + (\psi_2 - \psi_1) = 0 \quad (\text{Ec.10})$$

La versión discreta de la ecuación del flujo óptico es más frecuentemente usada para estimación de movimiento en video digital. Resolviendo la anterior ecuación para todos los \mathbf{x} se puede volver al problema de minimización con la siguiente función objetiva:

$$E_{OF}(\vec{a}) = \sum_{\mathbf{x} \in \Lambda} \left| (\nabla \psi_1(\vec{x}))^T \vec{d}(\vec{x}; \vec{a}) + \psi_2(\vec{x}) - \psi_1(\vec{x}) \right|^p \quad (\text{Ec.11})$$

Asignando $p = 2$, y suponiendo que el movimiento es constante en una pequeña región Λ' en la cual $\vec{d}(\vec{x}, \vec{a}) = \vec{d}_0$, siguiendo el desarrollo descrito se obtiene:

$$\vec{d}_0 = \left(\sum_{\vec{x} \in \Lambda'} \nabla \psi_1(\vec{x}) (\nabla \psi_1(\vec{x}))^T \right)^{-1} \left(\sum_{\vec{x} \in \Lambda'} (\psi_1(\vec{x}) - \psi_2(\vec{x})) \nabla \psi_1(\vec{x}) \right) \quad (\text{Ec.12})$$

Una ventaja de este método es que la función de minimización es una función cuadrática de los vectores de movimiento, cuando $p = 2$. Si los parámetros que describen el movimiento están linealmente relacionados con los vectores, entonces la función tiene un único mínimo y puede ser resuelta fácilmente. Esto no es cierto teniendo en cuenta el error DFD definido en la Ec.1. Sin embargo, la ecuación del flujo óptico es solo valida cuando el movimiento es pequeño, o cuando una estimación inicial del movimiento $\vec{d}'(\vec{x})$ puede ser calculada y se actualiza $\psi_2(\vec{x})$ a $\psi_2(\vec{x} + \vec{d}'(\vec{x}))$. Cuando la premisa no es cierta, es mejor utilizar el criterio del error DFD y buscar la solución mínima usando el gradiente descendente o el método exhaustivo de búsqueda.

Normalización:

Minimizando el error DFD o resolviendo la ecuación del flujo óptico no siempre se consigue una estimación de movimiento positiva. Esto se debe en parte porque la

presunción de intensidad constante no es siempre correcta. La intensidad de luz en el mismo punto de una imagen puede variar por efectos de la reflexión o de sombra. La segunda razón es que en una región con textura plana, muchas estimaciones de movimiento diferentes pueden satisfacer la presunción de intensidad constante o incluso la ecuación del flujo óptico.

Para obtener una solución física positiva, se deben imponer restricciones adicionales para normalizar el problema. Una propuesta de normalización es añadir un término adicional a la función de error (Ec.1) o (Ec.5), que forzaría a la estimación de movimiento resultante a soportar las características del movimiento común. Una propiedad del movimiento es que normalmente varía muy sensible de píxel a píxel, exceptuando en los límites de los objetos. Para cumplir ese criterio de la sensibilidad en la variación se puede utilizar un término de error entre los vectores de movimiento de píxeles adyacentes, siendo:

$$E_s(\vec{a}) = \sum_{\vec{x} \in \Lambda} \sum_{\vec{y} \in N_x} \left\| \vec{d}(\vec{x}; \vec{a}) - \vec{d}(\vec{y}; \vec{a}) \right\|^2, \quad (\text{Ec.13})$$

donde $\|\cdot\|$ representa la norma y N_x representa el conjunto de píxeles adyacente a \mathbf{x} .

El criterio de minimización puede ser expresado como:

$$E = E_{DFD} + w_s E_s. \quad (\text{Ec.14})$$

El coeficiente w_s se elige de acuerdo con la importancia de la sensibilidad en la variación del movimiento relativa al error de predicción. Para impedir que se produzca un emborronamiento excesivo, se reduce el término en los límites de los objetos, lo que a su vez da pie a otro problema interesante, la detección de dichas fronteras.

2.2.2. Flujo óptico

El sistema de visión humano percibe el movimiento mediante la identificación de un punto concreto cuya posición varía temporalmente. Esta identificación se realiza suponiendo que dicho punto no varía en color o brillo en su movimiento. Es interesante entender que el movimiento observado puede ser diferente al movimiento real en determinadas circunstancias, como ilustra el siguiente ejemplo.

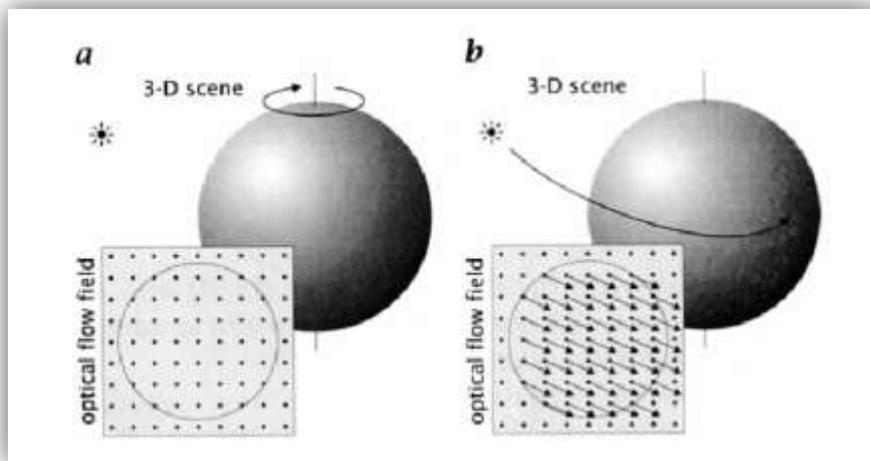


Figura 4: Ejemplo de las esferas. (a) Esfera girando con foco de luz estático (b) Esfera estática con foco de luz girando alrededor.

En el primer caso se tiene una esfera girando sobre su propio eje, en la que la luz reflejada es uniforme en toda la superficie visible. El observador puede tener la sensación de que al no haber diferencia de brillo, la esfera permanece estática. En el segundo caso se tiene una esfera en posición estática, y un foco de luz que gira en torno al eje de la misma. El observador tiene la sensación de que la esfera está girando sobre su propio eje.

De esta manera, uno es consciente de que el movimiento bidimensional real puede no corresponderse al flujo óptico observado, debido a que el flujo óptico no sólo depende del movimiento tridimensional, sino también de la iluminación y la textura de la superficie a observar. Cuando sólo se tiene la imagen de color, la mejor manera de estimar el movimiento es el flujo óptico. Sin embargo, en el resto

de los apartados se utilizarán los términos “movimiento” o “movimiento bidimensional” para hacer referencia al flujo óptico, siendo conscientes de que no siempre se corresponde el flujo óptico al movimiento real.

2.2.2.1. Ecuación del flujo óptico y ambigüedad en la estimación de movimiento

La ecuación del flujo óptico presenta unas características que provocan la necesidad de ajustes y la aplicación de **restricciones adicionales**, con el fin de poder resolverla. Se pueden enunciar tres características de la ecuación del flujo óptico.

1. No se puede determinar el vector de velocidad \mathbf{v} para cada uno de los píxeles de la imagen basándose únicamente en el gradiente espacial de la función de brillo $\nabla\psi$ y en la variación espacio temporal del mismo, $\partial\Psi/\partial t$. Existe una sola ecuación para dos incógnitas (v_x y v_y , o de manera análoga v_n y v_t). De hecho, la componente indeterminada es v_t , es decir, la componente perpendicular al gradiente de la función del brillo no puede ser determinada. Para hallar ambas incógnitas se deben imponer nuevas condiciones adicionales. La principal restricción es que los **cambios en el flujo óptico deben ser lentos**, no hay cambios bruscos en el movimiento entre puntos cercanos de la imagen, con lo que se puede hacer uso de la variación de intensidad en un espacio pequeño alrededor del punto \mathbf{x} para estimar el movimiento en el propio punto \mathbf{x} .
2. Dado el gradiente $\nabla\psi$ y la variación $\partial\Psi/\partial t$, la proyección del vector velocidad a lo largo de la dirección normal es fija, siendo $v_n = -(\partial\psi/\partial t)/\|\nabla\psi\|$, mientras que la proyección en la dirección tangencial es indeterminada, es decir, cualquier valor de v_t podría satisfacer la ecuación del flujo óptico. Esta ambigüedad se conoce como de **problema de apertura**.

“...the so-called ‘problem of aperture’ which also exists in human vision. This is related to the impossibility of unequivocally recovering the motion direction if

the object is observed through an aperture smaller than the object itself. In this context, the features on the object under observation (such as textures - patterns) are not enough to allow the perception of the transversal component of the object.” (32)

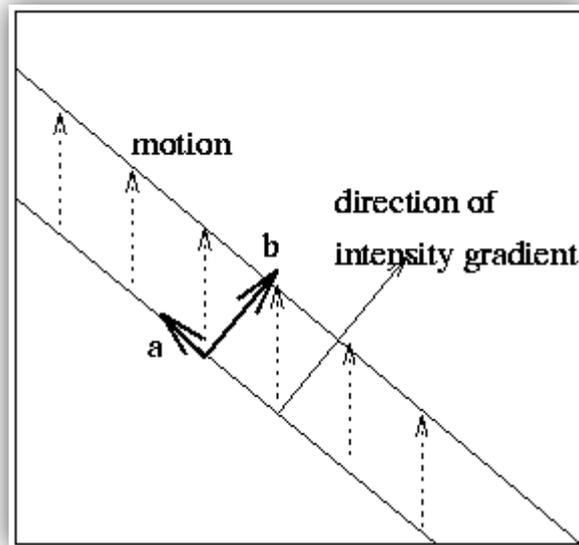


Figura 5: Problema de la apertura. Solo se puede determinar con absoluta certeza el vector b

Como se observa en el siguiente ejemplo, en el caso de una región uniforme, algunos cambios en la ventana de búsqueda nos pueden indicar no existencia de movimiento, cuando en realidad sí se experimenta un cambio espacial.

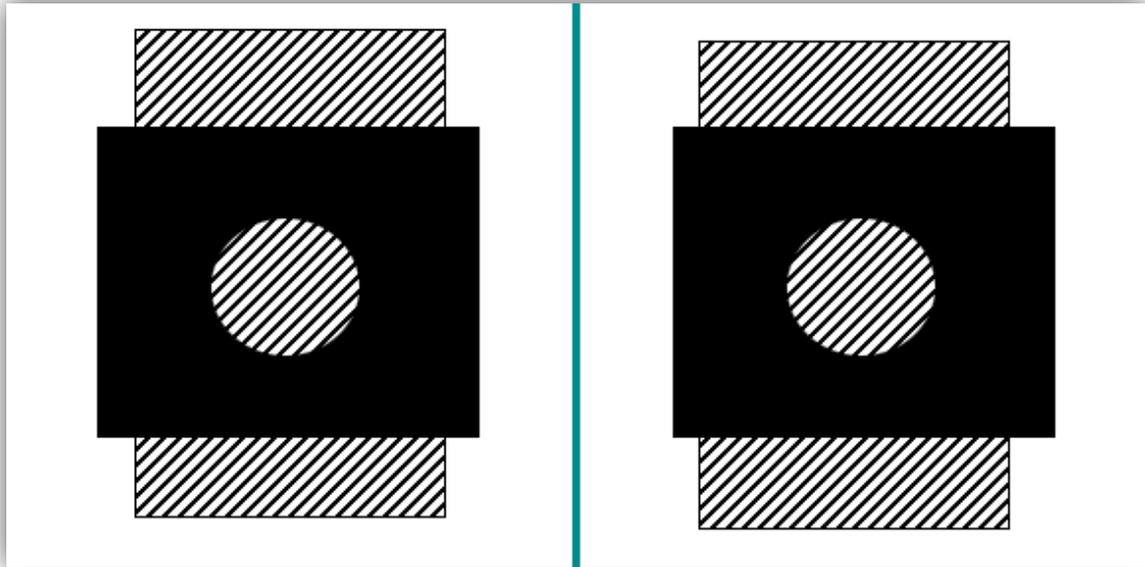


Figura 6: Moviendo la ventana de observación se pueden encontrar regiones exactamente iguales, que nos confunden a la hora de estimar el movimiento.

Para la resolución del problema de la apertura, se debe imponer una nueva restricción, que nos permita hallar la proyección en la dirección tangencial. Para ello, se impone **la restricción de región uniforme**, mediante la cual se supone que los píxeles colindantes a la región estudiada presentan el mismo patrón de movimiento. De esta manera, se resuelve el problema de la apertura, abriendo esa ventana estrecha que nos impedía observar el área adyacente.

3. En regiones con brillo constante, tales que $\|\nabla \psi\| = 0$, el vector del flujo es indeterminado. Esto se debe a que no se perciben cambios en el brillo cuando la superficie adyacente presenta un patrón uniforme. Es decir, **sólo se puede estimar el movimiento en las regiones con variaciones de brillo**, tales como bordes o texturas no uniformes.

Por otra parte, todas las características explicadas anteriormente presentan una serie de dificultades a la hora de su resolución matemática. Uno de los problemas más importantes es el estudio de secuencias en donde los objetos experimenten **grandes desplazamientos**. En este tipo de estudios, varios autores han empleado estrategias multiresolución (33) (34). En ellas se crea una pirámide de imágenes a

distinta resolución, estimando el movimiento desde el nivel de mayor resolución a niveles inferiores. De este modo se puede abordar desplazamientos mayores con un menor tiempo de procesado. Otro problema importante es la presencia de **oclusiones**. El estudio de oclusiones en los métodos basados en el gradiente presenta dos problemas. Por un lado, debido a la presencia de discontinuidades de movimiento, y por otro, debido a que en las áreas donde se presenta este fenómeno, la información de movimiento proporcionada por la ecuación del flujo óptico no será válida

Todo lo considerado en los apartados precedentes tienen un papel muy importante en los algoritmos de estimación de movimiento que se verán a continuación, con lo que es importante haber realizado dicho estudio del flujo óptico para tener una visión de los problemas que se nos plantean a continuación en lo relacionado con la estimación de movimiento.

2.2.3. Métodos basados en desplazamiento de bloques

Este método es uno de los métodos fundamentales en tratamiento de video, y está siendo utilizado por la mayoría de los estándares de compresión y codificación de video basados en la compensación de movimiento, que es la técnica que se encarga de la eliminación de redundancia temporal entre dos o más fotogramas sucesivos. Los algoritmos basados en desplazamiento de bloques se basan en realizar una división de cada una de las imágenes de una secuencia en bloques, que normalmente son rectangulares o cuadrados. El método consiste en detectar el movimiento entre las imágenes con respecto a los bloques que las constituyen. Para ello, se realizan una serie de operaciones de comparación entre cada uno de los bloques de la imagen previa con los bloques de un determinado entorno alrededor del mismo bloque en la imagen actual, la denominada región de exploración.

2.2.3.1. Región de exploración

Para realizar una estimación eficiente de los vectores de movimiento, se debe tener en cuenta que el tamaño de la región de exploración es crítico para la obtención de resultados. Dado que el coste computacional del algoritmo aumenta de forma casi cuadrática a medida que la ventana de búsqueda incrementa su tamaño, hay que conseguir un equilibrio entre su tamaño y el coste computacional soportable. También se tiene que identificar la aplicación del algoritmo para realizar la estimación de los parámetros tanto del tamaño de la ventana como el de la región. Por ejemplo, cuando se tiene un campo de exploración pequeño, los vectores de desplazamiento serán pequeños, lo cual es conveniente para movimientos lentos. En cambio, para objetos que experimentan movimientos rápidos, es conveniente la utilización de regiones de búsqueda más grandes, es decir, ventanas mayores. Se denomina longitud de paso de la búsqueda al número de píxeles que separa a los diferentes bloques candidatos que se van a analizar. Es decir, cuando dicha longitud es igual a la unidad, se está ante una *búsqueda exhaustiva*. En cambio, cuando la longitud de paso es mayor a la unidad se utiliza una *búsqueda rápida*, en la cual se incrementa la velocidad del proceso a cambio de incrementar el error de estimación.

2.2.3.2. Elección del tamaño bloque

Esta elección es clave en el proceso de búsqueda Block Matching, debido a que especifica la posición, el tamaño, la ubicación inicial y la escala de los bloques utilizados en el algoritmo.

Elegir el tamaño correcto de los bloques no resulta trivial. Generalmente los bloques mayores son menos sensibles al ruido, mientras que un bloque de dimensiones reducidas presenta unos contornos definidos. El principal factor a la hora de escoger las dimensiones de los bloques es el tamaño de los objetos a analizar, así como la presencia de ruido que presenta la secuencia o la textura de los objetos y el fondo de las imágenes. También existe el denominado *problema de apertura*, visto en el capítulo precedente, cuando se trabaja con objetos de color

uniforme. Los bloques en el interior de dichos objetos no parecen moverse porque los bloques colindantes poseen el mismo color. Para estos casos, el incremento del tamaño de bloque puede resultar eficaz.

Por otra parte, la mayoría de los algoritmos de Block Matching utilizan el origen de la ventana de exploración como el centro inicial de la búsqueda, sin explotar la correlación entre los bloques pertenecientes a un mismo objeto de la imagen en movimiento. Se puede implementar esta correlación para predecir una posición inicial que refleje la tendencia del movimiento del bloque, permitiendo encontrar el vector óptimo.

2.2.3.3. Criterios de comparación

Para explicar los diferentes criterios de semejanza, representados por $E(\mathbf{d}_m)$, se tomará un bloque cuadrado, B_m , de dimensiones $m \times m$. Se medirá la distancia entre un bloque del fotograma actual Ψ_2 y el fotograma previo Ψ_1 . El vector desplazamiento \mathbf{d}_m será el vector entre el bloque previo y el actual. Existen los siguientes criterios:

- *Suma de diferencias absolutas (Sum of Absolute Differences, SAD)*

$$E(\vec{d}_m, \forall m \in M) = \sum_{m \in M} \sum_{\vec{x} \in B_m} |\psi_2(\vec{x} + \vec{d}_m) - \psi_1(\vec{x})| \quad (\text{Ec.15})$$

- *Suma de diferencias al cuadrado (Sum of Squared Differences, SSD)*

$$E(\vec{d}_m, \forall m \in M) = \sum_{m \in M} \sum_{\vec{x} \in B_m} |\psi_2(\vec{x} + \vec{d}_m) - \psi_1(\vec{x})|^2 \quad (\text{Ec.16})$$

- *Error absoluto medio (Mean Absolute Difference, MAD)*

$$E(\vec{d}_m, \forall m \in M) = \frac{1}{n^2} \sum_{m \in M} \sum_{\vec{x} \in B_m} |\psi_2(\vec{x} + \vec{d}_m) - \psi_1(\vec{x})| \quad (\text{Ec.17})$$

- *Error cuadrático medio (Mean Squared Error, MSE)*

$$E(\vec{d}_m, \forall m \in M) = \frac{1}{n^2} \sum_{m \in M} \sum_{\vec{x} \in B_m} \left| \psi_2(\vec{x} + \vec{d}_m) - \psi_1(\vec{x}) \right|^2 \quad (\text{Ec.18})$$

- *PSNR (Peak Signal-to-Noise Ratio)*

También se debe definir el factor señal ruido, que es el valor que nos arrojará la precisión con la que se ha realizado la estimación de movimiento. Este valor calcula el error de la imagen compensada tras el proceso de estimación de movimiento con respecto a la imagen previa.

$$PSNR = 10 \log_{10} \left[\frac{X_{pic}}{MAD} \right], \quad (\text{Ec.19})$$

siendo X_{pic} el valor pico a pico de la señal original.

2.2.3.4. Métodos de búsqueda

Dependiendo de las características del sistema y de las propiedades de la secuencia a analizar, en ocasiones se debe anteponer el tiempo de cálculo a la exhaustividad de los resultados. Por ello, existen diferentes métodos para el cálculo de los vectores de movimiento por el método de *Block Matching*. El primero de ellos es el método exhaustivo (*Exhaustive Block Matching, EBM*), algoritmo que realiza una búsqueda del bloque óptimo por medio de la comparación del bloque con todos los bloques que se hallan dentro de la región de búsqueda de la imagen actual. A cambio de la precisión requiere un coste computacional muy alto. El resto de algoritmos de *Block Matching* buscan una compensación entre la precisión y el coste computacional, y para ello realizan una reducción del número de cálculos en el proceso, con el fin de obtener buenos resultados en un tiempo menor. El algoritmo de búsqueda en tres pasos (*Three Step Search, TSS*) fue uno de los primeros métodos diseñados para un cálculo más eficiente en la estimación de movimiento. Se basa en realizar comparaciones recursivas en las cuales se va reduciendo el tamaño de la región de búsqueda en cada uno de los tres pasos que conforman el método. Existen dos algoritmos más, que en realidad son modificaciones o mejoras del algoritmo TSS, el nuevo TSS (*New TSS, NTSS*), y el *Simple and Efficient TSS Search (SES)*. Ambos métodos realizan los cálculos en tres pasos, introduciendo una serie de mejoras en los procesos de elección de bloques

candidatos y del proceso de comparación. También existen métodos en cuatro pasos (*Four Step Search, FSS*), y métodos en los que la forma de la región de búsqueda presenta una distribución en forma de rombo (*Diamond Search*). Sumados a los métodos presentados también existen métodos en los que se realiza una búsqueda en espiral, o en los que se realiza una búsqueda cruzada de los bloques de la imagen actual. Todos ellos presentan diferentes características y son utilizados en situaciones concretas dependiendo de los requisitos de los sistemas. En este proyecto se analizarán y se implementarán los métodos que se cree que son más útiles y precisos, y se realizará una comparación entre los mismos.

2.2.4. Métodos basados en características

Los métodos basados en características extraen puntos de interés de la imagen, tales como líneas, contornos o regiones, mediante las que se determina un campo de movimiento. Para la estimación de movimiento se analiza el cambio de posición de dichos puntos característicos a través de varias imágenes consecutivas. Para realizar la estimación se hace corresponder las características extraídas de la imagen de referencia con las características encontradas en el resto de imágenes de la secuencia. Este esquema es usualmente utilizado con características simples, como puntos o líneas, debido a que la información que aporta cada característica es muy reducida, con lo que el rango de búsqueda se limita a una pequeña vecindad.

Desde finales de los años setenta se han ido desarrollando una serie de métodos basados en estos principios. Barnard y Thompson (35) (36), presentaron uno de los primeros y más conocidos métodos basados en correspondencia entre los puntos de interés encontrados en dos imágenes consecutivas. Su método se basaba en un procedimiento iterativo que calcula las probabilidades de correspondencia entre las dos imágenes. El cálculo se basa en la similitud local de los puntos y en el empleo de heurísticas que permiten reducir el espacio de búsqueda: máxima velocidad, coherencia de movimiento entre puntos y correspondencia biyectiva.

Otros métodos (37) utilizan los píxeles característicos extraídos mediante el método de Canny (38) para encontrar el flujo óptico de los bordes. Posteriormente se enfocó el problema de la correspondencia desde el punto de vista geométrico (39). El estudio se basa en considerar que los parámetros de movimiento obtenidos considerando todos los puntos de interés de dos imágenes permanecerán casi idénticos si se repite el análisis eliminando un par de puntos correspondientes.

En general, los métodos basados en características simples, tales como puntos y líneas, no resultan adecuados para la segmentación que se quiere obtener, una segmentación basada en movimiento. Esto es debido a que en su cálculo se producen una serie de campos dispersos de vectores de movimiento que nos dificulta la extrapolación para crear campos densos, sobre los cuales realizar el estudio que desemboca en la segmentación mejorada debido a la información que otorgan los vectores de movimiento.

Algunos autores (7) (40) consideran que este método puede ser a su vez dividido en dos métodos de búsqueda, que en cierta manera utilizan una correspondencia basada en características. Son los métodos de correspondencia de bloques (o desplazamiento de bloques) y los métodos basados en correlación de fase. Respecto al primer método, que se ha estudiado en profundidad en este mismo capítulo, se considera independiente debido a su popularidad y su importancia en el cálculo de la estimación de movimiento, pese a que en cierta manera sí mantiene cierta relación, en los fundamentos básicos, con los métodos basados en características.

Los métodos de correlación de fase (7) (40) estiman el desplazamiento entre dos imágenes por medio de una función de correlación cruzada normalizada, la cual es calculada en el dominio espacial bidimensional de Fourier. La principal idea es el principio por el cual una translación en el dominio espacial produce una variación en el término de la fase del dominio de Fourier.

2.3. MODELOS DE CLUSTERING

Los métodos de agrupamiento, o *clustering*, son procedimientos de agrupación de datos, generalmente vectores, de acuerdo con unas características comunes, como pueden ser la posición, en sistemas bidimensionales, o sistemas de magnitudes, como pueden ser las encontradas en el sistema realizado en el proyecto, en el que cada vector utilizado presenta dos magnitudes (dirección del vector) y dos variables que nos indican su posición (ejes). El proceso se realiza agrupando los datos según un criterio de cercanía, definida en términos de una determinada función de distancia, como la euclídea. Cada uno de los grupos, denominados *clusters*, comparte unas características comunes, lo que en este proyecto se traduce como que los distintos vectores de movimiento con características comunes pueden ser agrupados por regiones. Este hecho permite la representación del comportamiento de un grupo mediante un representante único, que es la región obtenida.

Existen una serie de algoritmos de clasificación. Uno de los más utilizados es el famoso *K-Means*, utilizado para la segmentación, para el reconocimiento de palabras y otras muchas áreas en el procesamiento de señales. También existen otros métodos probabilísticos de agrupamiento que miden la probabilidad de que un píxel determinado pertenezca a una burbuja (ya sea bidimensional o multidimensional) Gaussiana, y con el resultado obtenido se usan criterios de decisión. Sin embargo, para realizar el proceso, se utilizará el algoritmo *Mean Shift*, introducido por Fukunaga y Hostetler (41) y posteriormente revisado y actualizado por Dorin Comaniciu (42), que es un método no paramétrico utilizado para el análisis de espacios de características con densidades complejas multimodales.

3

DISEÑO

En el presente capítulo se realizará una descripción exhaustiva de los métodos que se han implementado en el Proyecto Fin de Carrera. Todos los métodos que se verán en el capítulo han sido enunciados y clasificados en el capítulo precedente, pero deben ser analizados en profundidad para conseguir un completo entendimiento de los objetivos del estudio.

Con el fin de obtener unos resultados precisos, en el diseño se realizarán una serie de comparaciones entre los métodos propuestos para cada uno de los modelos de los tres bloques principales, segmentación, estimación de movimiento y clustering y se implementará la arquitectura en base a todos y cada uno de ellos. De esta manera, se obtendrá un sistema complejo en el que analizar las secuencias de imágenes que se estudian en el proyecto.

La organización del capítulo será la siguiente:

1. **Modelos de segmentación seleccionados:** Para la segmentación, se opta por utilizar métodos de representación paramétricos, que aunque requieran un coste computacional más elevado ofrecen muy buenos resultados de

segmentación. Estos métodos caracterizan el modelo con una serie de parámetros que tienen como objetivo definir la distribución de cada píxel.

- ***Gaussiana Simple (SG)***: Método de representación paramétrico en el cual se representa cada uno de los píxeles con una distribución unimodal Gaussiana definida por la media y la varianza de la distribución.

2. ***Modelos de estimación de movimiento seleccionados***: Para el caso de la estimación de movimiento se estudiarán los dos métodos principales, ya sea la estimación por medio del flujo óptico y la estimación por medio de los métodos basados en el desplazamiento de bloques, que engloba los algoritmos de *Block Matching* existentes en el estado del arte.

- ***Métodos basados en el flujo óptico***: Método basado en la resolución de la ecuación del flujo óptico con la imposición de una serie de restricciones fundamentales. Se obtiene un vector de movimiento para todos y cada uno de los píxeles de la imagen.
- ***Métodos Block-Matching***: Métodos basados en el desplazamiento de bloques. Se realizará un estudio de los seis métodos que se han considerado más precisos.

3. ***Modelos de clustering seleccionados***: Para realizar el agrupamiento se implementará uno de los métodos de *clustering* existentes en el estado del arte.

- ***Algoritmo Mean Shift***: Técnica no paramétrica de análisis de características, que nos realizará la etapa de agrupamiento de los vectores con características comunes.

3.1. MODELO DE SEGMENTACIÓN IMPLEMENTADO

Para realizar la segmentación de una secuencia de imágenes se debe escoger un método de segmentación que cumpla unos requisitos de calidad determinados. Se necesita un método de segmentación con el que se obtengan resultados fiables y precisos logrando un equilibrio entre la calidad y la carga computacional.

Dados todos los métodos de segmentación enunciados en capítulos anteriores, se decide tomar como modelo de estudio la segmentación basada la Gaussiana Simple (GS), basándonos en los tiempos de procesamiento y en la calidad mostrada en las pruebas realizadas a lo largo del Proyecto Fin de Carrera de Sonsoles Herrero Martín (13). Este método presenta unos resultados muy precisos y posee una estructura sencilla y simple para su implementación. Otro de los puntos que hacen que sea un método de interés para el actual proyecto es que los resultados que se obtienen dependen únicamente de tres parámetros, lo que hace que sea cómodo a la hora de modificarlos para las distintas secuencias.

Como se ha visto anteriormente, los modelos paramétricos son aquellos que realizan una estimación del valor de cada píxel del fondo de las secuencias por medio de funciones de distribución de probabilidad. Estas distribuciones deben actualizarse con cada nueva imagen de entrada. En vez de utilizar un buffer con el que estimar el fondo, los píxeles del fondo se representan por los parámetros de la distribución. La distribución matemática que se utilizará para modelar el fondo será la distribución Gaussiana $N(\mu, \sigma^2)$. Existen varios métodos que utilizan esta distribución con el fin de discernir qué píxeles de la imagen pertenecen al fondo y cuales al frente. Dado que el principal objetivo es realizar una mejora de la segmentación inicial, se debe escoger uno de los métodos que mejores resultados ofrece. Este método es el basado en la Gaussiana Simple, método en el cual cada píxel del fondo se representa con dos parámetros , (μ, σ^2) utilizados para el proceso de selección de los píxeles.

3.1.1. Gaussiana Simple (SG)

El método basado SG consiste en un modelado de fondo por medio de una Gaussiana Simple. Se denomina Gaussiana simple en referencia a que toda la distribución se realiza con la misma función de probabilidad. Mediante este método, se realiza el ajuste de la densidad de probabilidad de una función Gaussiana (función de densidad de probabilidad, p.d.f. de sus siglas en inglés), en cada uno de los píxeles de la imagen a procesar. El método consiste en que para cada píxel del modelo de fondo $B_t(x,y)$ se mantiene el valor de luminancia medio $\mu_t(x,y)$ y una desviación estándar $\sigma_t(x,y)$. De aquí en adelante se utilizará la misma notación.

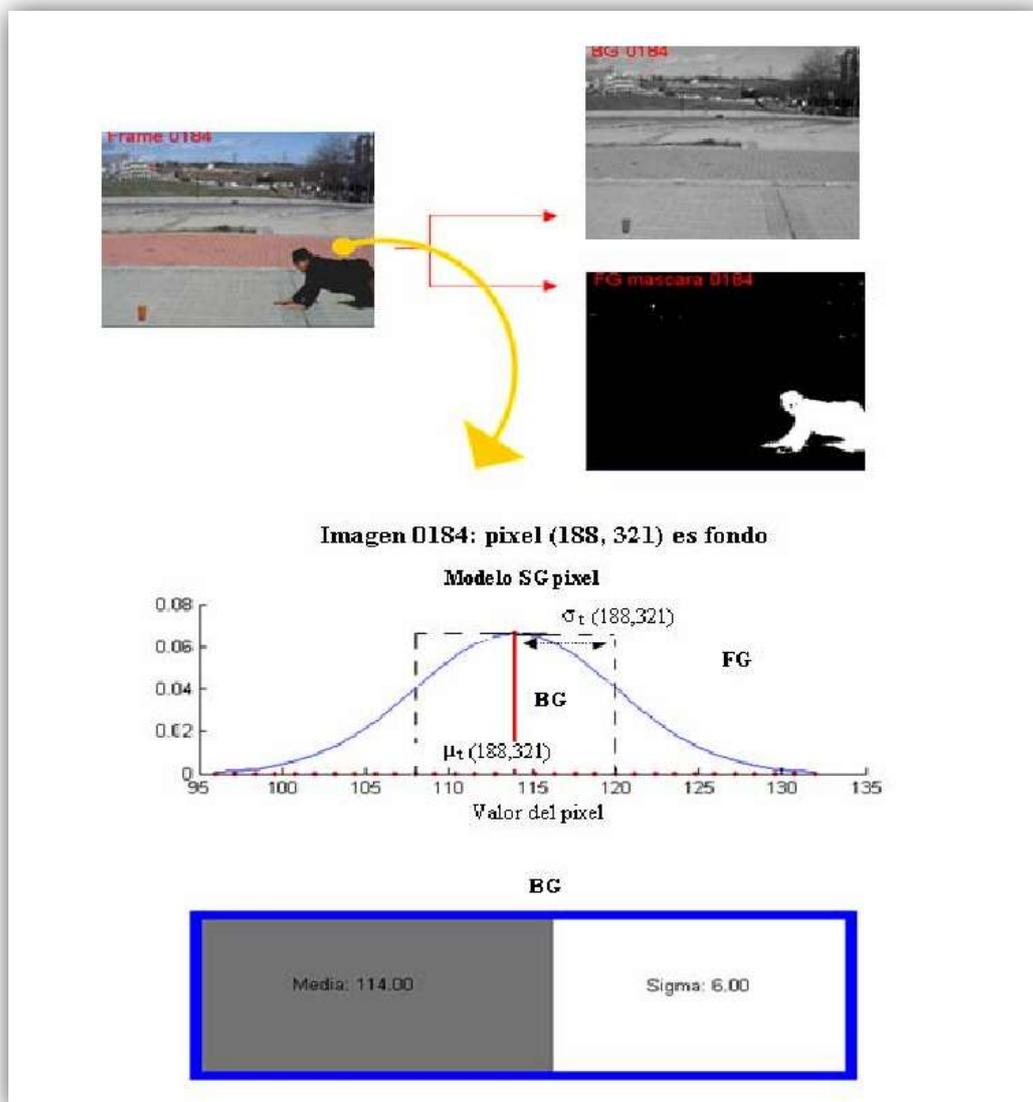


Figura 7: Distribución Gaussiana del píxel (188, 321) de una secuencia ejemplo en el fotograma 184.

En el proceso de inicialización se creará un modelo del fondo primario, que se irá actualizando posteriormente, y que proporcionará valores a los parámetros que serán utilizados como valores de referencia. De esta manera se compara cada uno de los píxeles de la nueva imagen con los valores de media que nos indican el fondo de la imagen. Para saber si un píxel pertenece al fondo o al frente se utiliza un criterio de decisión basado en que si la diferencia entre ambos valores supera k veces la desviación establecida para cada píxel (el valor de k suele ser de 2 o 3 desviaciones) el píxel a estudio forma parte del objeto en movimiento.

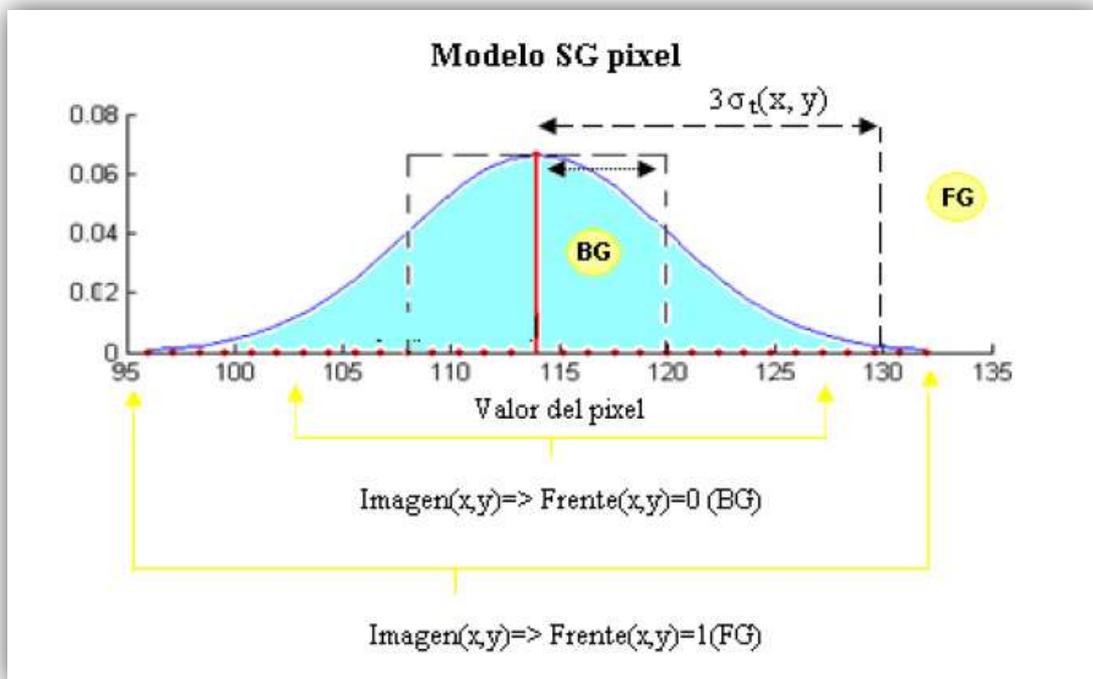


Figura 8: Distribución Gaussiana de un píxel

Para la implementación del algoritmo que crea la máscara final, se deben realizar una serie de etapas. Son las siguientes:

- **Inicialización del fondo y parámetros**

El método inicializa la imagen la imagen de fondo a la primera imagen de la secuencia y como valor de desviación se marcan todos los píxeles a un valor

pequeño que se denominará P . Este valor de desviación varía en función de las características de la secuencia. Si su fondo contiene mucho movimiento, el valor de la desviación será elevado (valores entre 25 o 35), para permitir modelar los píxeles con una Gaussiana más ancha y permitir más rango de variación en los valores que los píxeles pueden tomar en las imágenes de la secuencia. Por otra parte, se tendrá un valor pequeño P (de entre 5 a 10) para ajustar el fondo a la secuencia y poder decidir con mayor precisión qué píxeles forman parte del objeto y cuáles no.

- **Estimación del frente**

Para discernir cual de los píxeles pertenecen al fondo y cuales al frente, se debe calcular la posibilidad de que el píxel de la nueva imagen procesada (I_t) pertenezca a la Gaussiana que modela el fondo, y cuyos parámetros se han calculado en el paso anterior. Una función Gaussiana modela una densidad de probabilidad que deja de tener valor a aproximadamente dos o tres veces el valor de la desviación de la función. Por esto, si la diferencia entre la imagen y el modelo de fondo (la media) del píxel difiere en $k=2-3$ veces el valor de la desviación, se considera que no hay parecido entre el nuevo píxel y su modelo, y se determina que el píxel es objeto en movimiento (F_t). Si sucede lo contrario será marcado como fondo.

$$|I_t(x, y) - \mu_t(x, y)| \geq k\sigma_t(x, y) \Rightarrow F_t(x, y) = 1$$

$$|I_t(x, y) - \mu_t(x, y)| < k\sigma_t(x, y) \Rightarrow F_t(x, y) = 0$$

- **Actualización del fondo**

Los parámetros que modelan el fondo de la secuencia se actualizan por la utilización de una media móvil caracterizada por la velocidad de cambio de la secuencia ($\alpha \ll 1$). Este parámetro es fundamental en la obtención de resultados, y será relevante en los apartados siguientes del proyecto.

$$B_{t+1}(x, y) / \begin{cases} \mu_{t+1}(x, y) = \alpha I_t(x, y) + (1 - \alpha)\mu_t(x, y) \\ \sigma_t^2(x, y) = \alpha(I_t(x, y) - \mu_t(x, y))^2 + (1 - \alpha)\sigma_t^2(x, y) \end{cases}$$

Este modelo, cuenta con una serie de ventajas, que han hecho que finalmente sea utilizado. Es un método sencillo, rápido en su ejecución, se adapta automáticamente al ruido y a los cambios progresivos del fondo. A su vez, tiene una serie de desventajas, como puede ser que no soporte fondos multimodales en los que los píxeles oscilan entre valores muy distintos, como puede ser cierto parpadeo en la luminancia de las hojas de los árboles, o luces intermitentes. Además, cada secuencia posee un valor muy diferente de varianza (σ^2) y velocidad de cambio (α), que deben ser optimizados para ofrecer los mejores resultados en todas las situaciones estudiadas.

3.2. MODELOS DE ESTIMACIÓN DE MOVIMIENTO IMPLEMENTADOS

El proceso de estimación de movimiento tiene una gran importancia en este proyecto, debido a que la mejora de la segmentación que se desea conseguir depende en gran manera de los resultados que se obtengan en el proceso de estimación de movimiento. Precisamente por este motivo, se debe realizar un examen exhaustivo de los métodos presentes en el estado del arte, decidiendo finalmente con cuál de los métodos que se implementan se obtienen los resultados más precisos.

En esta etapa, lo que interesa fundamentalmente es la calidad de los vectores de movimiento obtenidos, sin tener en cuenta la alta carga computacional que determinados métodos puedan tener, debido a que ésta es la etapa crítica del proyecto, y los resultados obtenidos al estimar el movimiento serán la base de todo estudio que se realizará posteriormente. Para ello, se estudiarán los métodos basados en el flujo óptico y los métodos de *Block Matching*.

Se debe indicar que los métodos basados en características no han sido evaluados, debido a que ofrecen poca precisión en los casos que se quieren determinar. Esto se debe a que los métodos basados en características estudian el movimiento en determinados puntos de la imagen (puntos característicos) y en base a su movimiento, determinan el movimiento de los objetos presentes en la escena. En el caso que se trata en este proyecto, lo que interesa es determinar los máximos vectores de movimiento posibles, para tener conciencia del movimiento en todos los puntos de la imagen.

3.2.1. Método basado en el flujo óptico

En el método basado en el flujo óptico (o gradiente) que se ha implementado se intenta estimar un vector de movimiento para cada uno de los píxeles de la imagen. Teniendo en cuenta la restricción de intensidad constante, para cada punto de la imagen previa habrá muchos píxeles en la imagen actual que tengan el mismo valor de intensidad. Utilizando la ecuación del flujo óptico que se ha enunciado en el capítulo anterior, el sistema sigue siendo indeterminado debido a que existe una ecuación para dos incógnitas. Para solucionar el problema se deben imponer **cuatro restricciones** en la implementación de la ecuación del flujo óptico:

1. La primera consiste en utilizar técnicas de normalización para hacer cumplir la **restricción de movimiento lento** en el campo, para que los vectores de movimiento que aparecen nuevos en una imagen (debido a que el píxel no existía en la imagen precedente), estén determinados o definidos por los vectores de los píxeles adyacentes.
2. En segundo lugar, se debe asumir que los vectores de movimiento en una región alrededor de un punto determinado son iguales, y aplicar la **restricción de intensidad constante** o la ecuación del flujo óptico sobre la región completa, no sólo sobre los píxeles que la forman.
3. La tercera es una restricción adicional de invarianza. Además de la restricción de intensidad constante, que nos lleva a la ecuación del flujo

óptico, se puede asumir que el **gradiente de la intensidad es constante al movimiento**, como se propone en (43) (44) (45).

4. Y por último se debe hacer uso de la **relación entre las funciones de fase** entre la imagen previa y la actual (40).

A continuación, se enunciarán las dos primeras restricciones que se han utilizado para la implementación del algoritmo de estimación de movimiento basado en el flujo óptico. Es importante su enunciado, debido a que es de gran utilidad en la consecución de resultados correctos en el proceso de estimación, y el motivo por el cual lo enuncio en este apartado es que son los procesos concretos en los que se basa el algoritmo desarrollado.

3.2.1.1. Normalización mediante la restricción de movimiento lento

El método para la estimación de vectores de movimiento propuesto por Horn y Schunck (46) consiste en la minimización de la función que resulta de combinar la ecuación de flujo óptico y el criterio de suavidad en el movimiento:

$$E(\vec{v}(\vec{x})) = \sum_{\vec{x} \in \Lambda} \left(\frac{\partial \psi}{\partial x} v_x + \frac{\partial \psi}{\partial y} v_y + \frac{\partial \psi}{\partial t} \right)^2 + w_s (\|\nabla v_x\|^2 + \|\nabla v_y\|^2) \quad (\text{Ec.20})$$

Originalmente, los gradientes espaciales de v_x y v_y son aproximados por:

$$\begin{aligned} \nabla v_x &= [v_x(x, y) - v_x(x-1, y), v_x(x, y) - v_x(x, y-1)]^T \\ \nabla v_y &= [v_y(x, y) - v_y(x-1, y), v_y(x, y) - v_y(x, y-1)]^T \end{aligned}$$

La minimización de esta función de error se lleva a cabo mediante el método basado en el gradiente, conocido como el método de Gauss-Siedel.

3.2.1.2. Restricción de región uniforme

En este método, cuando se estima el vector de movimiento del punto \mathbf{x}_n , se asume que todos los vectores de movimiento de los píxeles alrededor de \mathbf{x}_n pertenecientes a la región $\beta(\mathbf{x}_n)$ son iguales, y los son denominados \mathbf{d}_n . Para determinar \mathbf{d}_n se puede minimizar el error de predicción sobre $\beta(\mathbf{x}_n)$ o por el contrario se puede resolver la ecuación el flujo óptico usando el método de mínimos cuadrados. En el desarrollo se utilizará la primera aproximación, estimar \mathbf{d}_n mediante el error DFD sobre la región $\beta(\mathbf{x}_n)$.

$$E(\vec{d}_n) = \frac{1}{2} \sum_{\vec{x} \in \beta(\vec{x}_n)} w(\vec{x}) (\psi_2(\vec{x} + \vec{d}_n) - \psi_1(\vec{x}))^2, \quad (\text{Ec.21})$$

donde $w(\mathbf{x})$ son los pesos asignados al píxel \mathbf{x} . El gradiente con respecto a \mathbf{d}_n será:

$$\bar{g}(d_n) = \frac{\partial E}{\partial d_n} = \sum_{\vec{x} \in \beta(\vec{x}_n)} w(\vec{x}) e(\vec{x}, \vec{d}_n) \frac{\partial \psi_2}{\partial \vec{x}} \Big|_{\vec{x} + \vec{d}_n}, \quad (\text{Ec.22})$$

donde $e(\vec{x}, \vec{d}_n) = \psi_2(\vec{x} + \vec{d}_n) - \psi_1(\vec{x})$ es la diferencia de la imagen desplazada (DFD) de \mathbf{x} con el \mathbf{d}_n estimado. Denominando $\vec{d}_n^{(l)}$ a la estimación en la iteración l -ésima; el gradiente de primer orden daría lugar a:

$$\vec{d}_n^{(l+1)} = \vec{d}_n^{(l)} - \alpha \bar{g}(\vec{d}_n^{(l)}). \quad (\text{Ec.23})$$

Partiendo de la ecuación Ec.13, la actualización de cada iteración depende de la suma de los gradientes de la imagen en varios píxeles, escalados por el peso de los valores de la DFD en dichos píxeles.

En lugar de utilizar el algoritmo de actualización basado en el gradiente, se puede realizar una búsqueda exhaustiva para encontrar el \mathbf{d}_n que da lugar al mínimo error con un determinado margen de búsqueda. Lo que nos ofrece una conexión con el siguiente método a explicar, y que será parte fundamental en el desarrollo del proyecto. La diferencia con respecto al método explicado consiste en que en la búsqueda exhaustiva se utiliza una ventana deslizante, y se determina un vector de

movimiento para cada píxel, mediante un proceso de minimización del error en su mismo entorno o región.

3.2.1.3. Estrategia multiresolución

En el presente estudio, se ha apreciado que existen determinadas situaciones en las cuales el movimiento resulta demasiado grande para poder ser calculado de manera precisa u otras en el que las regiones son uniformes y no nos permiten el cálculo preciso del movimiento que se produce realmente. De este modo, el cálculo de los desplazamientos puede hallar mínimos locales en las regiones analizadas, debido precisamente a que el punto se ha desplazado fuera de la región de búsqueda. Para alcanzar el mínimo global de la búsqueda, existen estrategias basadas en la multiresolución (47): se utilizan distintas resoluciones en la imagen utilizada.

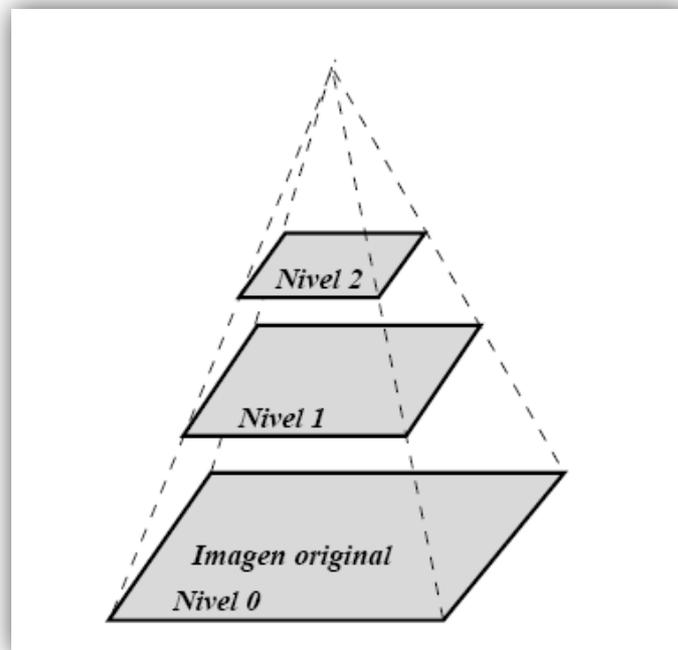


Figura 9: Pirámide de imágenes. Cada nivel corresponde a una resolución diferente.

El método implementado se basa en un algoritmo recursivo de búsqueda multiresolución, en el que se busca el desplazamiento que sufre cada uno de los píxeles de la imagen a diferentes resoluciones. Comenzando con una resolución más burda, se identifica el desplazamiento de los puntos de la misma. Esta solución primaria sirve como inicio para la resolución de la versión refinada de la misma imagen. Con lo que paso a paso se puede resolver el problema del desplazamiento más rápido que puede producirse en determinados objetos de la imagen. Es decir, se usará un método que iterativamente va mejorando la estimación, donde en cada iteración se selecciona aquella estimación con error mínimo.

En el método implementado aparecen una serie de variables que determinan la calidad del sistema.

- ***Alpha global***: Parámetro que define la variable global de movimiento lento. Mediante este término, se asegura que exista flujo óptico en regiones sin textura, regiones homogéneas donde el movimiento puede resultar imposible de obtener, así como en regiones en las cuales existe algún tipo de oclusión. En definitiva este parámetro se asegura del cumplimiento de la restricción de movimiento lento, enunciada anteriormente.
- ***Alpha local***: Este parámetro especifica el criterio de movimiento lento pero de un modo dependiente del gradiente, es decir, en las regiones de la imagen donde existen esquinas o límites de objetos, preservando de esta manera las discontinuidades cerca de los bordes, que normalmente se deben al movimiento de los objetos o a oclusiones. De esta manera se asegura el cumplimiento de la restricción de movimiento lento en zonas locales de la imagen.
- ***Número de niveles de resolución***: Este valor se calcula en dependencia con el tamaño de las imágenes que se quieren analizar, y queda determinado por la siguiente ecuación:

$$num_levels = round(\log_{10}(30 / \min(ht, wt)) / \log_{10}(0.9))$$

En donde el valor 30 se refiere al mínimo tamaño posible del ancho o largo de la imagen en su menor nivel de resolución y los valores de ht y wt son el alto y el ancho de la imagen analizada (generalmente 576 y 720 respectivamente). El valor asignado de 0.9 hace referencia a la proporción entre tamaños de las imágenes de niveles contiguos de la imagen laplaciana creada en el proceso. Estos valores se asignan a priori, debido a que resultan óptimos en las pruebas que se han realizado pese a presentar un coste computacional elevado.

3.2.2. Métodos basados en desplazamiento de bloques implementados

Como se ha visto anteriormente, el principal problema del método basado en el gradiente es que para normalizar el problema se debe imponer la restricción de intensidad constante. Un método para imponer dicha restricción en la estimación de movimiento es dividir la imagen en pequeñas regiones no solapadas, denominadas bloques, y suponer que el movimiento para cada uno de ellos puede ser caracterizado por un modelo paramétrico, ya sea constante, afín, o bilineal. Cuanto más pequeño sea el bloque más precisa será la aproximación.

En teoría, los bloques pueden tener cualquier forma poligonal, pero en la práctica se utilizan normalmente bloques cuadrados o triangulares. A lo largo del estudio se implementarán los algoritmos basándonos en bloques cuadrados y a continuación se describirán algoritmos de estimación de movimiento basados en búsqueda por bloques, en donde el fin de los mismos será encontrar un vector de movimiento para cada uno de los bloques.

El criterio de semejanza determina la elección del bloque con mayor similitud, es decir, el que minimiza un error medido, de entre los candidatos dentro de la ventana de búsqueda de tamaño fijo del fotograma previo. Si el bloque elegido no se encuentra en la misma posición en ambas imágenes, significa que ha cambiado de posición. La distancia del bloque coincidente entre el fotograma actual y el previo se define como el **vector de desplazamiento estimado**, y será el que se le

asigne a todos los píxeles del bloque. En el caso ideal, los píxeles correspondientes de los bloques coincidentes serían exactamente iguales. No obstante, ese caso sucede en muy raras ocasiones, ya que la forma de los objetos en movimiento varía con respecto al punto de vista del observador o la luz reflejada sobre su superficie, y los resultados siempre se verán afectados por el ruido.

En el caso de que bloques conexos presenten el mismo patrón de desplazamiento, se pueden realizar combinaciones entre ellos, para formar regiones más amplias, muy útiles para diversas aplicaciones, métodos que se verán en el siguiente capítulo para obtener máscaras de segmentación basadas en regiones.

A continuación se realizará una explicación práctica de los algoritmos de estimación de movimiento que utilizan el principio de desplazamiento de bloques para la estimación. Los siguientes algoritmos se basan en la búsqueda de bloques colindantes que presentan mayor similitud con el actual. Dependiendo de su algoritmo de búsqueda, se puede alcanzar una estimación de movimiento con mayor o menor número de pasos, a cambio de obtener una menor o mayor exactitud de búsqueda. (48)

3.2.2.1. Algoritmo Block-Matching exhaustivo (EBMA)

El método exhaustivo determina el vector desplazamiento óptimo \mathbf{d}_m para un bloque \mathbf{B}_m dado del fotograma previo por medio de la comparación con todos los bloques candidatos en el fotograma actual. Esta comparación se realiza determinando una región de búsqueda y encontrando el bloque para el cual el error es mínimo. El vector de movimiento estimado sería el desplazamiento entre ambos bloques. Para este método se utilizará el error absoluto medio (MAD). La región de búsqueda es normalmente simétrica con respecto al bloque actual, con R_x píxeles a los lados y R_y píxeles tanto arriba como abajo. La precisión viene determinada por el paso de búsqueda, que es la distancia entre los bloques candidatos más cercanos, tanto en horizontal como en vertical. En el proyecto se utilizarán bloques cuadrados, en donde $R_x = R_y = R$ (denominado desde ahora como

máximo desplazamiento) así como un paso igual en ambas direcciones igual a la unidad, lo que normalmente se conoce como *integer-pel accuracy search*.

Teniendo un bloque de $N \times N$ píxeles, y un rango de búsqueda de $\pm R$ píxeles tanto en el sentido horizontal como vertical. Con un paso de un píxel, el número total de bloques candidatos para cada bloque en la imagen previa es $(2R + 1)^2$. El número de operaciones necesarias para calcular el MAD para cada candidato estimado es de N^2 , con lo que para cada bloque necesitaré $(2R + 1)^2 N^2$ operaciones. Por lo cual para una imagen de tamaño $M \times M$ píxeles, se tienen $(M/N)^2$ bloques (asumiendo que N es múltiplo de M). Con estas premisas, para una imagen determinada, se deben realizar $M^2(2R + 1)^2$ operaciones. Es muy importante resaltar que el número de operaciones totales no depende del tamaño de bloque, del valor de N .

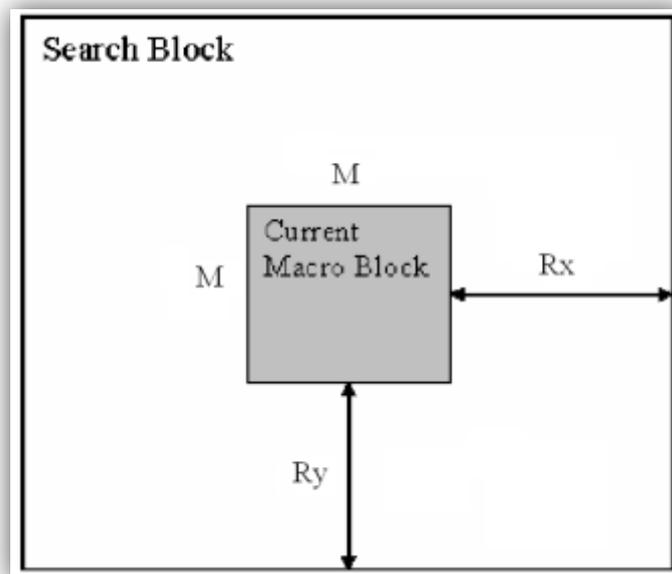


Figura 10: Sistema de búsqueda con bloque de tamaño $M \times N$

Ejemplo:

Para una imagen de 512×16 píxeles, y con un valor de $R = 16$, el número total de operaciones por imagen es de 2.85×10^8 . Con lo que para una secuencia con 30 fps,

el número se incrementa a 8.55×10^9 operaciones, lo que resulta una cifra astronómica.

De esta manera se observa cómo el método exhaustivo requiere una carga computacional muy alta, que puede impedir que sea utilizado en aplicaciones en tiempo real. Por otra parte, una ventaja del EBMA es que puede ser implementada por hardware, utilizando un diseño simple y modular, y pudiendo ser acelerado utilizando módulos en paralelo.

Algoritmos rápidos

El método exhaustivo realiza una gran cantidad de operaciones en su cálculo, lo que supone una muy alta carga computacional. Con el fin de acelerar el proceso, se han desarrollado una serie de métodos en los que se antepone la velocidad a la precisión. La clave para disminuir la carga es reducir el número de bloques candidatos. Los siguientes métodos rápidos difieren en el modo de elección de los candidatos y sin embargo arrojan errores pequeños.

3.2.2.2. Three Step Search (TSS)

El algoritmo en tres pasos fue uno de los primeros algoritmos desarrollados para una estimación rápida de movimiento, en la que el coste computacional se veía disminuido considerablemente respecto a las búsquedas exhaustivas definidas con anterioridad, y aportó simplicidad y alto rendimiento a los métodos de estimación. El algoritmo parte con la definición de un paso de búsqueda (S) y de un punto central. Posteriormente se compara el bloque a estimar de la imagen previa con los bloques que forman un cuadrado alrededor del mismo punto en la imagen actual. De esta manera se tiene el punto central y ocho puntos alrededor a una distancia entre sí de un paso de búsqueda. El punto que represente al menor error medido, siguiendo cualquiera de los métodos de comparación explicados en el apartado precedente, será el centro de la siguiente iteración. A continuación el paso se divide entre dos ($S' = S/2$) y se realizan las mismas comparaciones. El algoritmo

El método de búsqueda TSS utiliza un patrón de búsqueda uniformemente repartido y es propenso a no tener en cuenta movimientos pequeños. En la primera iteración, se comprueban 16 puntos, para obtener el punto en el que el error se minimiza. Estos 16 puntos están formados por el punto central, 8 puntos alrededor y separados una distancia predefinida como paso de búsqueda (por ejemplo, $S=4$) y los otros 8 puntos están situados alrededor del punto inicial con $S = 1$. Posteriormente se obtiene el menor peso de los 16 puntos a comparar. En el caso de que el punto de menor coste se encuentre en los puntos externos, el algoritmo se resolverá como el caso TSS. Si el menor coste está en los puntos internos, se realiza una comparación entre los 3 o 5 puntos alrededor del mínimo coste no comparados anteriormente. La naturaleza de los 3 o 5 puntos reside en la posición del punto en el cuadrado, como se observa en la siguiente figura.

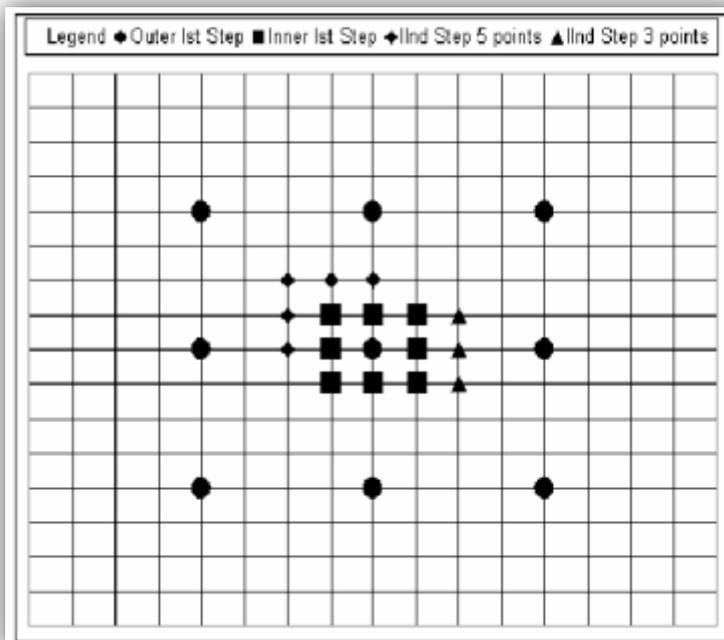


Figura 12: Búsqueda NTSS. Puntos grandes pertenecen a primera iteración. Triángulos son los puntos de búsqueda añadidos en NTSS. Los triángulos y los diamantes son el segundo paso de búsqueda, dependiendo de que sean 3 (triángulos) o 5 (diamantes)

El algoritmo compara en el mejor de los casos 17 posiciones, mientras que realiza comparaciones con 33 puntos en el caso peor.

3.2.2.4. Simple and Efficient TSS Search (SES)

El método SES (50) es también una extensión del método TSS. La idea principal es que para una superficie unimodal no puede haber dos mínimos en direcciones opuestas, con lo cual los 8 puntos fijados en los métodos anteriores para las búsquedas en tres pasos pueden cambiarse para incorporar esta máxima y así reducir el número de operaciones.

El algoritmo se divide en tres pasos, pero la novedad es que cada paso está formado por dos fases. El área de búsqueda está dividida en cuatro cuadrantes y el algoritmo comprueba tres posiciones A, B, y C, como se muestra en la siguiente figura. A está en el origen y B y C se encuentran a un paso de $S = 4$ en direcciones ortogonales. Dependiendo de los pesos que se obtengan en las comparaciones, se seleccionan los puntos pertenecientes al centro de las iteraciones posteriores FIGURA 13. Las reglas para determinar el cuadrante de búsqueda en las fases siguientes son:

Si $MAD(A) \geq MAD(B)$ y $MAD(A) \geq MAD(C)$, se selecciona el modelo (b).

Si $MAD(A) \geq MAD(B)$ y $MAD(A) \leq MAD(C)$, se selecciona el modelo (c).

Si $MAD(A) < MAD(B)$ y $MAD(A) < MAD(C)$, se selecciona el modelo (d).

Si $MAD(A) < MAD(B)$ y $MAD(A) \geq MAD(C)$ se selecciona el modelo (e).

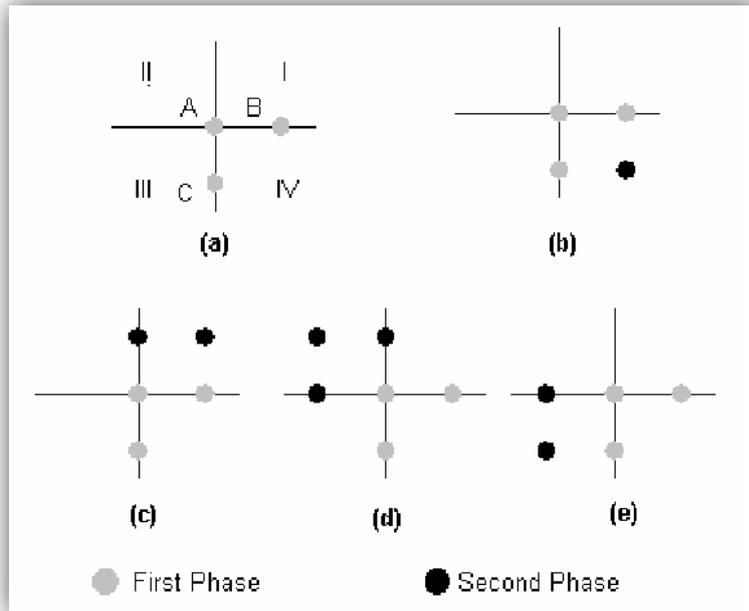


Figura 13: Patrones de búsqueda correspondientes a cada cuadrante:
 (a) Todos los cuadrantes. (b) Cuadrante I. (c) Cuadrante II.
 (d) Cuadrante III. (e) Cuadrante IV.

Una vez se hayan seleccionado los puntos de la segunda fase, se comparan los puntos para encontrar el lugar en el que el error es menor. Posteriormente se cambia el paso de la misma manera que en TSS y se repite recursivamente hasta alcanzar un paso de búsqueda igual a la unidad ($S = 1$).

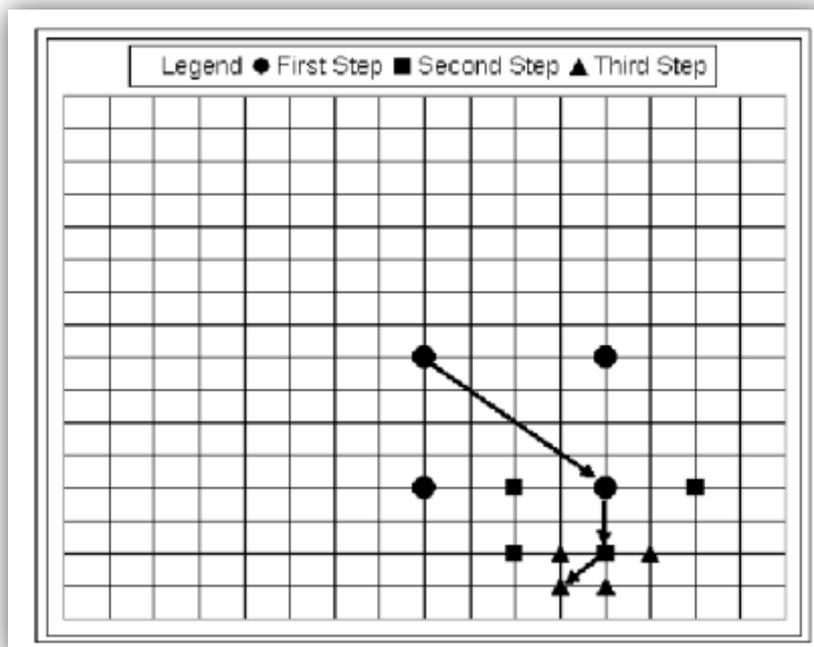


Figura 14: Algoritmo SES. Vector resultado (3,-7)

A pesar de que este método ahorra muchos cálculos con respecto a TSS, no es aceptado ampliamente por dos razones. La primera es que el error superficial no es estrictamente unimodal, con lo que la PSNR lograda es pobre, comparada con TSS. La segunda razón es que existe otro algoritmo, denominado Four Step Search (FSS), que fue presentado un año antes y ofrece menos coste computacional que TSS y además una tasa de error PSNR considerablemente mejor.

3.2.2.5. Four Step Search (4SS)

Siendo parecido al algoritmo SES, el 4SS (51) también emplea una búsqueda parcial y tiene la posibilidad de saltarse pasos intermedios en el caso de que se cumpla alguna condición. 4SS fija un patrón de tamaño de paso de $S = 2$ en el primer paso, independientemente del valor asignado a R . Se realizan comparaciones con el punto central y los 8 puntos alrededor del centro, en las direcciones cardinales, teniendo una ventana de búsqueda de 5×5 . En el caso de que el mínimo error se encuentre con el píxel del centro, el algoritmo salta al cuarto paso de búsqueda, que consiste en una búsqueda con paso de $S = 1$ en torno al punto central. Cuando el mínimo error se encuentra en los 8 puntos exteriores, se realiza la misma búsqueda, centrándonos en el punto más parecido y con un paso de búsqueda de $S = 2$. Ahora se tendría que comparar el punto óptimo con los 3 o 5 puntos alrededor no comparados anteriormente (Figura 14). Esta fase vuelve a tener la posibilidad de ser acelerada cuando el punto óptimo es el central, y se pasaría directamente al paso cuarto. Exactamente lo mismo ocurre en el tercer paso, del que obligatoriamente se salta al cuarto y último estado del algoritmo.

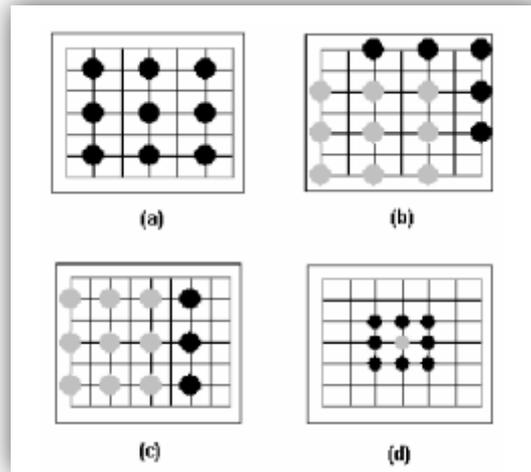


Figura 15: Patrones de búsqueda en 4SS. (a) Primer paso. (b) Segundo paso. (c) Tercer paso. (d) Cuarto paso.

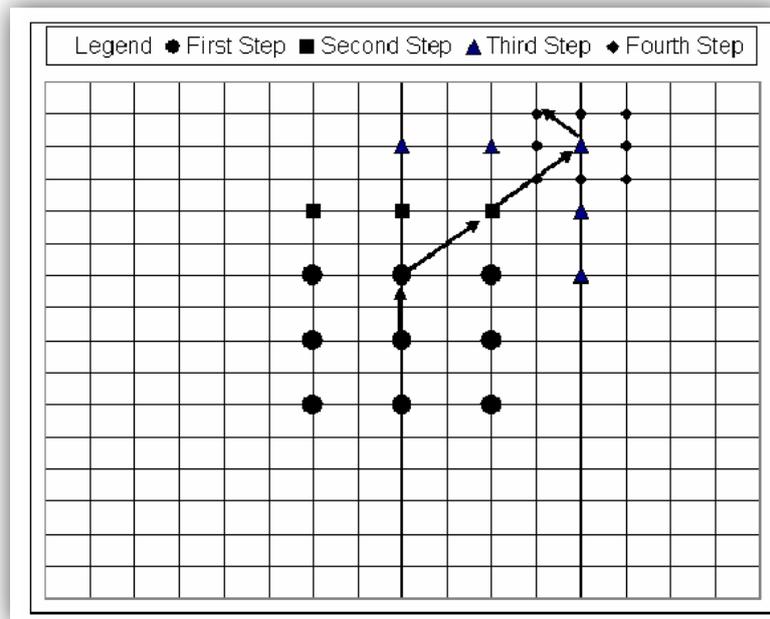


Figura 16: Procedimiento de búsqueda 4SS. Punto óptimo (3,7)

Este método tiene en el mejor de los casos 17 puntos de búsqueda, y 27 en el peor.

3.2.2.6. Diamond Search method

El método basado en búsqueda por diamante (52) es exactamente el mismo que la el método basado en la búsqueda con cuatro pasos, pero el patrón de búsqueda cuadrado se cambia a un patrón con forma de rombo, y además no hay límite en el

numero de pasos que el algoritmo puede realizar. Este método utiliza dos patrones de búsqueda, uno grande, Large Diamond Search Pattern (LDSP), y uno más pequeño, Small Diamond Search Pattern (SDSP), como se puede observar en la figura mostrada a continuación. El método de búsqueda es similar que para el método en cuatro pasos: se realiza una primera comparación mediante el patrón grande, y si el punto de menos peso está en el centro, se inspeccionan los puntos mediante el patrón pequeño. Mientras el punto óptimo no se encuentre, se realizan los mismos pasos que en el método de cuatro etapas, utilizando tres o cinco nuevos puntos para formar los rombos de búsqueda. Como este método no tiene número de pasos limitados, los resultados ofrecidos pueden llegar a ser muy precisos, alcanzando una PSNR muy próxima en valor a la alcanzada mediante el método exhaustivo de búsqueda, mientras que el coste computacional es significativamente menor.

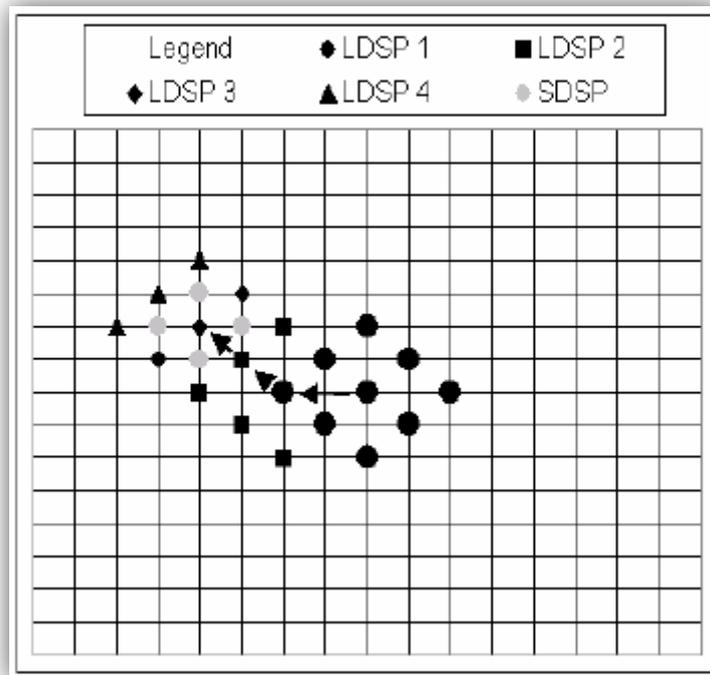


Figura 17: Método Diamond Search.
 Este ejemplo muestra un punto óptico en $(-4,-2)$ alcanzado en la quinta iteración.

3.3. MÉTODO DE CLUSTERING IMPLEMENTADO

El algoritmo de *Mean Shift* es una técnica general no paramétrica que se utiliza para encontrar grupos en espacios de características. Éste método fue desarrollado por D. Comaniciu y P. Meer en el año 2002 (42). La utilización de esta técnica se basa en una aplicación iterativa del método para encontrar el máximo local o punto estacionario de la función de densidad más cercano a un punto P del conjunto de datos. La principal ventaja que ha hecho que se implemente este método y no otros es que para su ejecución no es necesario el conocimiento a priori del número de grupos, y además permite la forma arbitraria de éstos.

El motivo para utilizar una técnica general no paramétrica de estimación de la densidad se debe a que el espacio de características se puede considerar como la función de densidad del parámetro representado. Regiones densas en el espacio de características se corresponden con los máximos locales de la función de densidad de probabilidad, es decir, las modas de la densidad desconocida. Una vez se ha encontrado la moda para cada punto, se asociarán en un grupo común todas aquellas que se encuentren en un contorno determinado.

3.3.1. Algoritmo Mean Shift

El algoritmo de agrupamiento basado en Mean Shift consiste en la aplicación práctica del procedimiento de búsqueda de modas, según se describe a continuación.

- Para los puntos del conjunto de datos mantener el procedimiento de búsqueda de puntos estacionarios de $f_{K,h}(x)$ o modas.
- Deshacerse de los puntos intermedios reteniendo sólo el máximo local.

En otras palabras, el algoritmo Mean Shift agrupa un conjunto de datos de dimensión asociando a cada punto con la moda o el pico de la función de densidad de probabilidad del conjunto de datos. Para cada punto, el algoritmo calcula la

moda correspondiente y la asocia a él. En primer lugar, se define un kernel radial centrado en el punto y con ancho de banda h y se calcula la media de los puntos que caen dentro del área definida bajo el kernel. Después, el algoritmo desplaza la ventana o el kernel hacia la media y se va repitiendo el proceso hasta que converge, es decir, hasta que el mean shift vector calculado recursivamente se mantiene constante o hasta que su variación es inferior a un umbral determinado. En cada iteración, la ventana se desplazará hacia zonas de mayor densidad de puntos hasta que se alcance el pico, donde los datos están distribuidos de la misma manera en la ventana. El algoritmo siguiente muestra el pseudocódigo del agrupamiento Mean Shift. Al finalizar el algoritmo, todos los puntos asociados a una moda de valor similar son asignados a un mismo grupo.

Algoritmo de agrupación Mean Shift

Definiendo:

m_h = Mean Shift Vector.

Entrada: Conjunto de datos X donde $x_i \in R^d$, $i = 1 \dots n$ y ancho de banda h .

```

1: for all  $x_i \in X$  do
2:    $x_t \leftarrow x_i$ 
3:   while  $m_h(x_t) \neq \text{umbral}$  do
4:     Calcular  $m_h(t)$ 
5:      $x_t \leftarrow x_t + m_h(t)$ 
6:   end while
7:    $\text{ClustCent}_i = \{x \in D \mid \text{dist}(x, x_t) \text{ es mínima}\}$ 
8:   if  $\text{ClustCent}_i$  no existe then
9:      $\text{Cluster}_i = \text{nuevo ClusterId}$ 
10:  end if
11:end for

```

Salida: Vector $\text{cluster} \in R^d$.

La idea es seleccionar una región de interés y pasarla por una ventana. En la región de interés seleccionada, se busca el valor medio de todos los puntos de su interior. Y se le asigna el nuevo centro de masas al punto obtenido:

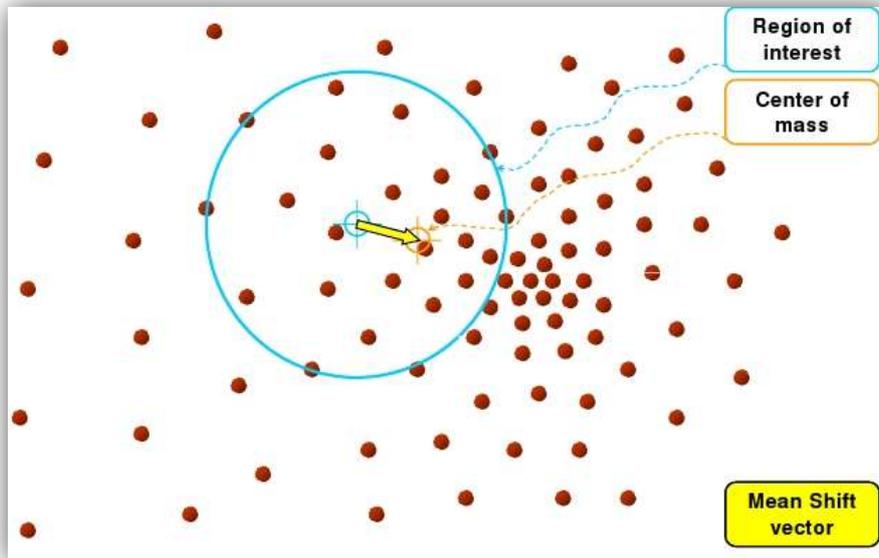


Figura 18: Fase inicial del algoritmo

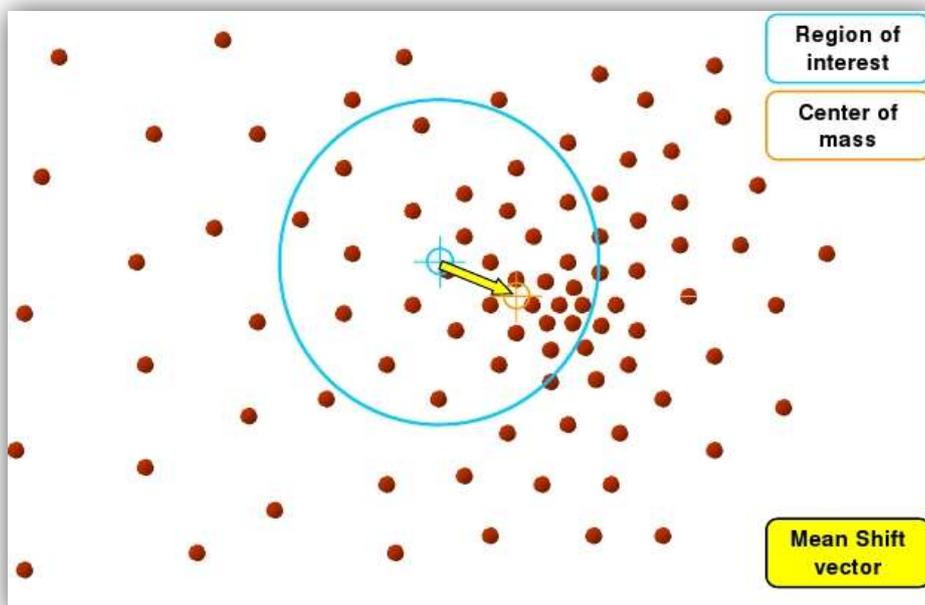


Figura 19: Segunda iteración. Se busca el punto óptimo.

El algoritmo continúa buscando hasta que se alcance una convergencia en el área de interés. Para ello, se debe suponer que todo píxel que nos dirija hacia el mismo máximo local de la convergencia forma parte de la misma región. Para ello, se debe

realizar un proceso previo de búsqueda de máximos locales, o modos, que existen en la imagen. Por último se realiza el etiquetado, que significa asignar a cada punto de la imagen el modo correspondiente.

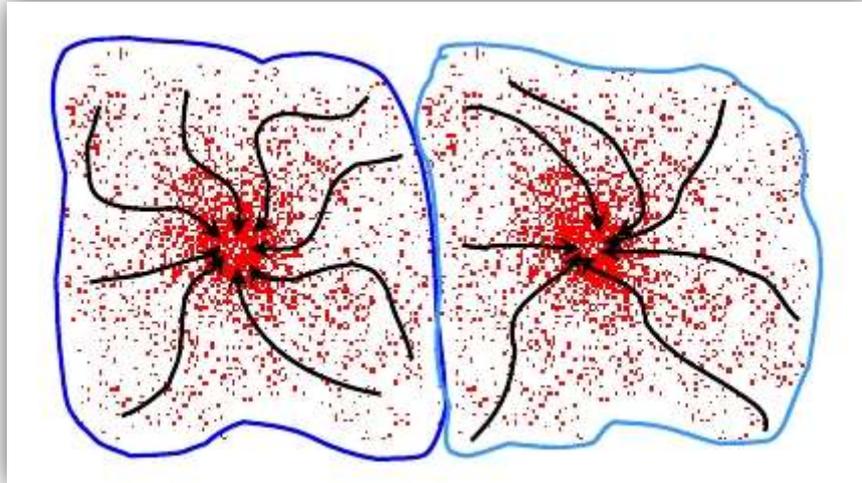


Figura 20: Proceso finalizado con dos regiones encontradas

4

ALGORITMO PROPUESTO

En este capítulo se describe la arquitectura que se ha realizado para la comprobación de los modelos y obtención de resultados. Es una arquitectura basada en cuatro bloques principales, en la que de un modo automático e iterativo se consigue una mejora en la segmentación inicial por medio del estudio de la estimación de movimiento que se realiza a la secuencia de imágenes.

Como muestra el siguiente diagrama, el bloque inicial consiste en la **segmentación**. De este modo se obtiene una segmentación de partida para el sistema. Posteriormente se realizará una batería de pruebas, en las que se variarán los parámetros característicos que modificarán los resultados e influirán en la precisión adquirida. El bloque siguiente será una **estimación de movimiento**, también realizada por varios métodos explicados en el resto del proyecto, con lo que se tendrá otro abanico de resultados. Posteriormente se continúa con la fase de **clustering**, o agrupamiento, mediante la cual se agruparán los vectores de movimiento acorde con sus características. Para terminar, se realizará una **actualización de la segmentación inicial**, basada en los resultados obtenidos al

realizar el *clustering* y basándonos en un criterio iterativo de elección y de comprobación que será explicado ampliamente en el presente capítulo.

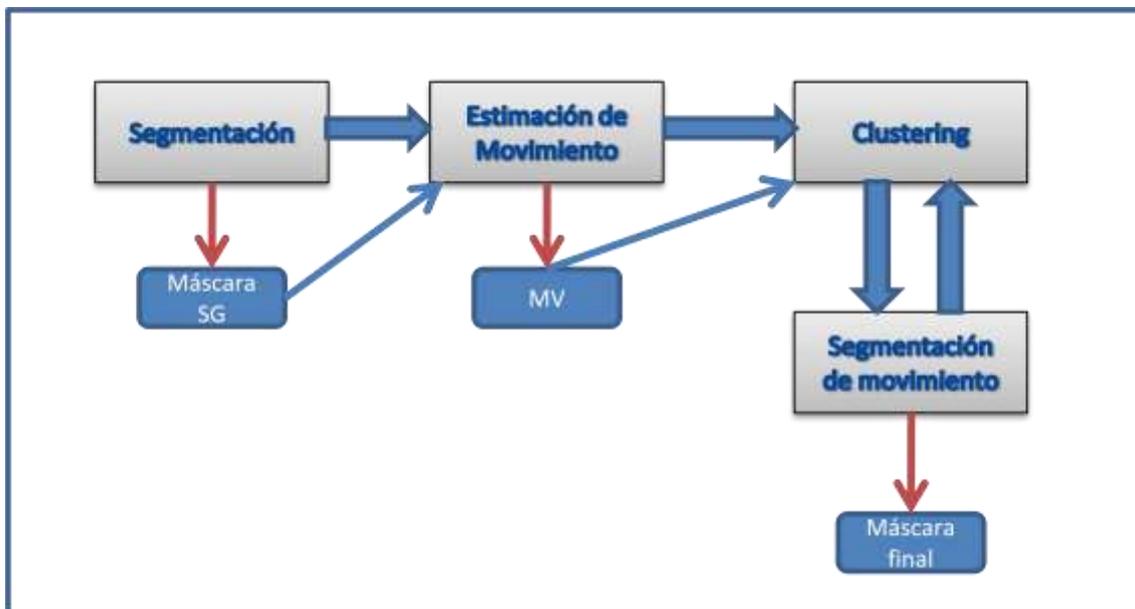


Figura 21: Diagrama de las etapas de la integración.

4.1. ETAPA DE SEGMENTACIÓN

Cabe destacar que en el proyecto se parte de una segmentación determinada, escogida con el fin de obtener unos resultados más precisos, pero que puede ser sustituida por cualquiera de las segmentaciones existentes en el estado del arte. Esto se debe a que la arquitectura puede ser modificada, cada uno de los bloques son independientes, y porque lo que se intenta realizar es una mejora de la segmentación, sea cual sea el método con la que se realice. Es suficiente, por lo tanto, obtener una segmentación ideal, que se utilizará basándonos en las máscaras *ground truth*, y una segmentación que aporte resultados fiables y precisos, en nuestro caso una segmentación basada en el modelado de fondo, a su vez extraído mediante el método basado en la Gaussiana Simple (SG). Este punto es importante, porque demuestra el sentido del proyecto: realizar una mejora de la segmentación inicial.

La segmentación mediante el método de la Gaussiana Simple tiene una serie de parámetros que serán utilizados para obtener la mayor precisión posible, dependiendo de las características de la imagen. Los parámetros a modificar que han sido estudiados en el apartado 3.1.1 serían los siguientes:

- **La desviación de la distribución.**
- **La media del modelo de fondo (alpha).**
- **El factor k .**

Tanto para los valores de la desviación y de la media, como para los valores del parámetro k se realizarán pruebas previas a la integración completa del sistema con el objetivo de establecer los valores de estos tres parámetros óptimos para cada secuencia de imágenes estudiada. El parámetro k es el valor referencia que decide si un punto de la imagen pertenece al fondo o al frente. Si la diferencia de valores de un píxel en las dos imágenes supera k veces la desviación establecida, el píxel forma parte del objeto en movimiento. El valor de k será modificado tomando valores de entre 1 y 4 veces la desviación establecida.

Para comenzar, se debe realizar la **lectura de las imágenes** a analizar. En el método propuesto se utilizará una secuencia de imágenes con un número no menor a 90 muestras, con el fin de obtener una segmentación precisa de los objetos que presentan movimiento a lo largo de la secuencia completa. Las imágenes que se leen presentan un formato de imagen '.png' y serán almacenadas en disco.

A continuación, se **inician los parámetros** del proceso de **segmentación**, conociéndose las características de las imágenes. Este punto es fundamental para la obtención de resultados fiables, debido a que las características de cada secuencia hacen variar los resultados de manera significativa. Este proceso creará una **máscara** de los puntos de la imagen que han sufrido movimiento que a su vez será almacenada en disco para poder ser estudiada en las siguientes fases del algoritmo.

Es importante haber obtenido previamente las máscaras *Ground Truth*, en el caso del proyecto han sido proporcionadas por miembros del *Video Processing and Understanding Lab*, de la Universidad Autónoma de Madrid (53), que serán las **máscaras ideales** de las imágenes a procesar. De este modo, se obtienen los resultados de la segmentación práctica (en esta caso mediante el método de la Gaussiana Simple), y de la segmentación ideal (*Ground Truth*), que serán utilizados para la comprobación de la precisión e idealidad del método final.



Figura 22: Ejemplo de máscara resultado de la segmentación basada en SG



Figura 23: Ejemplo de máscara Ground Truth.

En el ejemplo se observa cómo mediante la segmentación basada en la Gaussiana se obtienen zonas del fondo que en el *Ground Truth* no aparecen, así como zonas de la silueta de la persona que no deberían aparecer y no son detectadas. El fin del algoritmo propuesto es eliminar lo máximo posible las zonas exteriores a los objetos que aparecen en las secuencias y aumentar cuanto se pueda las zonas de siluetas propiamente dichas, que representan a los objetos presentes en el frente.

4.2. ETAPA DE ESTIMACIÓN DE MOVIMIENTO

Una vez extraídas las máscaras de movimiento mediante la segmentación basada en la Gaussiana Simple (SG), se debe estimar el movimiento de las imágenes a estudio. Para el análisis, únicamente es necesario tener dos imágenes, con lo que se seleccionan las dos últimas de las utilizadas para la segmentación. Una vez elegidas las imágenes, se debe **superponer la máscara obtenida con las imágenes**, de esta manera se obtienen dos imágenes en las que aparecen en blanco las regiones que no han experimentado movimiento, y la imagen real en las zonas en las que sí se ha experimentado algún tipo de movimiento. De esta manera se hallan los vectores de movimiento en las zonas que se considera que han experimentado un

movimiento. Para que el proceso de estimación de movimiento sea más preciso, ampliamos el área alrededor de las zonas donde se detecta movimiento con el fin de comparar también los puntos de los bordes de los objetos.

Posteriormente, se analiza el movimiento mediante los métodos seleccionados, previamente explicados en el apartado 3.2 del presente trabajo. Estos métodos son el basado en el gradiente, el **método del flujo óptico**, y las distintas versiones de los métodos basados en desplazamiento de bloques, como son los diversos algoritmos de **Block Matching**. Todos y cada uno de los métodos son realizados en las mismas imágenes obtenidas, y cada uno de ellos nos devolverá un resultado diferente, mostrando las virtudes y defectos de todos ellos.

En cuando a la **inicialización de los parámetros**, se debe diferenciar cada uno de los dos grupos. Para el caso de la estimación de movimiento basada en el flujo óptico, los parámetros son:

- **La media global de búsqueda (alpha global).**
- **La media local de búsqueda (alpha local).**
- **El numero de niveles para el proceso multiresolución.**

Mientras que para los métodos basados en el desplazamiento de bloques son:

- **El tamaño seleccionado para cada bloque.**
- **El máximo desplazamiento aceptado en cada búsqueda.**

Para todos los métodos presentados se realizará exactamente el mismo procedimiento. Se **estiman los vectores de movimiento** de las imágenes estudiadas, y como resultado, se obtiene una matriz que contiene los vectores de movimiento obtenidos en la estimación. Esta matriz, no tendrá las mismas dimensiones para cada método, depende de si los vectores son calculados para cada píxel, o para cada bloque, con lo que presentará unas dimensiones dependientes del tamaño de bloque escogido. Para posteriores estudios, se debe normalizar la medida de las matrices obtenidas.

En la siguiente figura, se observan los vectores de movimiento obtenidos para los fotogramas 199 y 200 de la secuencia BAILE. En la primera imagen se observan los contornos que se han obtenido tras la estimación, formados por los vectores de movimiento resultado del proceso basado en el flujo óptico.

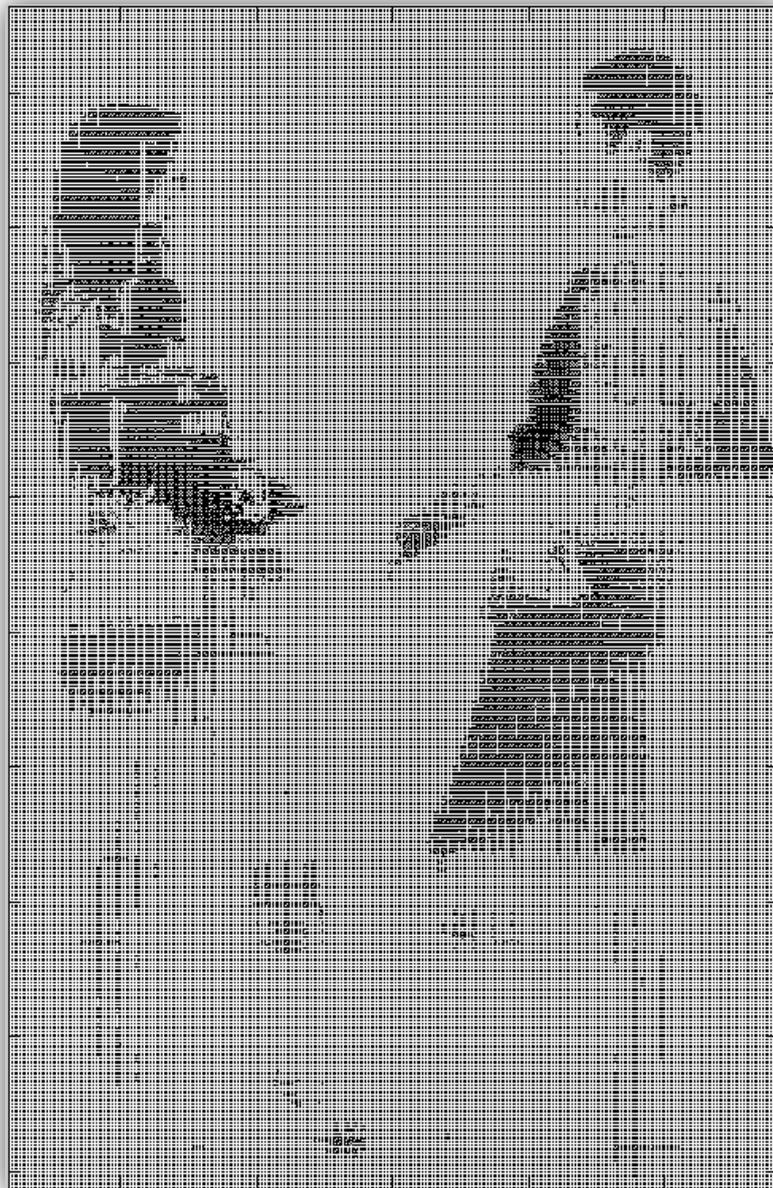


Figura 24: Estimación de movimiento para los fotogramas 199 y 200 de la secuencia BAILE

En la siguiente figura, observamos con mayor resolución una zona de la estimación del movimiento, concretamente la cabeza de una de las dos personas que conforman la secuencia. En ella, se observa la tendencia del movimiento que sufren los puntos de la imagen en esta región.



Figura 25: Región de la estimación de movimiento

4.3. ETAPA DE AGRUPAMIENTO O CLUSTERING

Realizada la etapa de estimación de movimiento, se realiza el proceso de agrupamiento. Para ello se hará uso del método de **Mean Shift** explicado en profundidad en el apartado 3.3. Dicho método calcula las agrupaciones de píxeles con unas características parecidas en un entorno determinado. Para obtener los resultados correctos, se debe otorgar un ancho de banda coherente con el movimiento que se esté estudiando. Por lo tanto, el proceso de **inicialización** del algoritmo Mean Shift, se basa en establecer los valores de:

- El **ancho de banda** requerido para calcular los **vectores similares**.
- El **ancho de banda** que realiza la comparación en un **área** determinada.

Dependiendo de las características de la imagen, se variarán estos dos parámetros, puesto que: el valor que limita el área de agrupamiento debe ser mayor para objetos con dimensiones también mayores, y el valor de ancho de banda para el cálculo de la similitud entre vectores hará que existan un mayor número de regiones, a cambio de una mayor precisión, o al revés.

En las imágenes mostradas a continuación se muestra una máscara de segmentación inicial para la secuencia INGRAVIDEZ, con un valor de **k** igual a 3. Posteriormente, tras haber realizado la estimación de movimiento, los resultados ofrecidos por el proceso de **Mean Shift** muestran las distintas regiones encontradas en la matriz de vectores de movimiento. Las zonas con mismo color pertenecen a regiones con vectores de movimiento coherentes entre sí.



Figura 26: Máscara de segmentación mediante SG con valor $k = 3$.

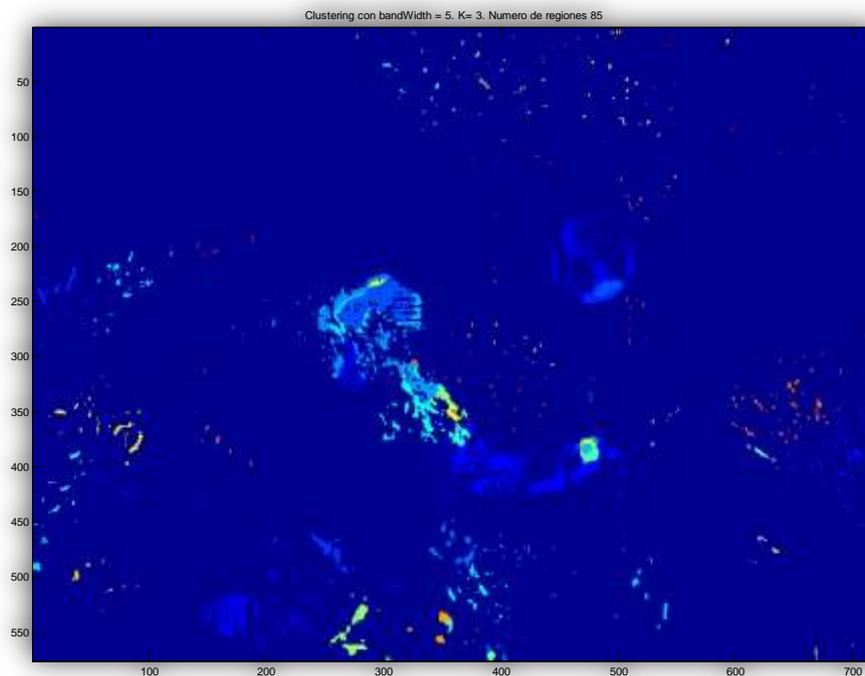


Figura 27: Resultado de Mean Shift sobre los vectores de movimiento de la imagen *enmascarada* con la máscara de la figura anterior.

Para comenzar el proceso de análisis del movimiento se realiza **clustering de todos los vectores** obtenidos, incluidos en las zonas en las que se detectó

movimiento para cada uno de los distintos valores de k que se escojan. De esta manera, se obtienen todas las regiones presentes en la matriz de vectores de movimiento calculada. Para todos los procesos, se otorgará un ancho de banda diferente, con el fin de obtener unos resultados óptimos. Se denomina **Bandwidth1**, a este primer valor del ancho de banda. En el siguiente diagrama observamos el proceso:

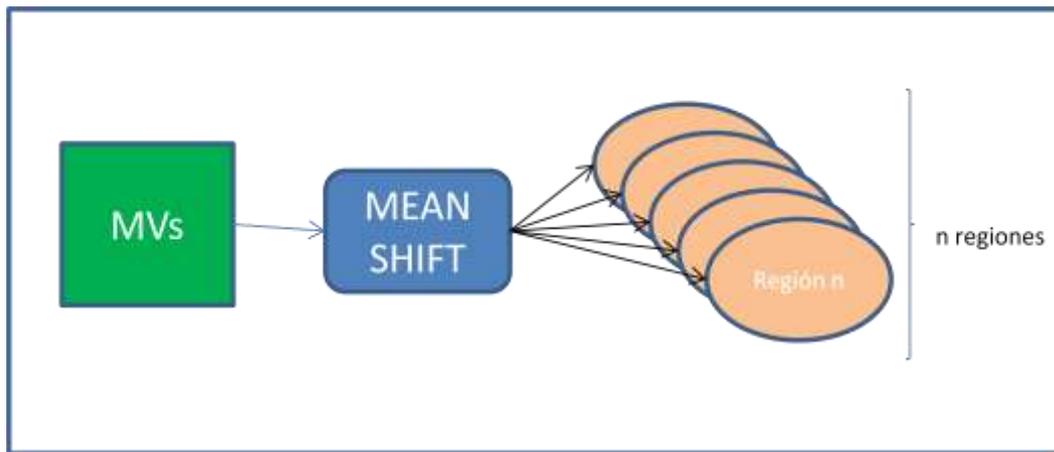


Figura 28: Diagrama del proceso de clustering.

El proceso de clustering se realiza de manera conjunta con el proceso que se ha denominado segmentación de movimiento. Para cada una de las iteraciones que se realizan en la última etapa del algoritmo, es necesario conocer la naturaleza de las regiones que conforman la imagen. Por ello, se volverá a utilizar el proceso de agrupamiento en numerosas ocasiones a lo largo de la siguiente etapa.

4.4. ETAPA DE SEGMENTACIÓN DE MOVIMIENTO

A lo largo de las pruebas realizadas durante la realización del proyecto, se ha observado que el factor principal que influye en mayor medida en los resultados de la segmentación es el parámetro k . Se llega a la conclusión que habiéndose obtenido los valores óptimos de la media y la varianza para cada secuencia de imágenes, el factor k determinará la calidad de la segmentación.

Como se observa en los ejemplos siguientes de segmentación obtenidos para la secuencia BAILE, para cada valor de k obtenemos una segmentación diferente. Para valores mayores de k el proceso de segmentación detecta menos zonas de movimiento en la secuencia que para los valores más bajos de k .



Figura 29: Segmentación mediante SG con $k = 2$.



Figura 30: Segmentación mediante SG con $k = 3$.



Figura 31: Segmentación mediante SG con $k = 4$.

4.4.1. Segmentación por movimiento inicial

Para cada una de las máscaras con un k distinto, se ha realizado el proceso de *Mean Shift* para los vectores de movimiento obtenidos, que arroja una representación de las regiones que presentan un movimiento común. Partiendo de los datos obtenidos en estos agrupamientos, se tienen tantas representaciones de las regiones como valores de k se hayan elegido. Estas representaciones son almacenadas en forma de matriz, en la que cada píxel tiene como valor el número de la región a la cual pertenece. Esto será útil posteriormente a la hora de la identificación de las regiones en las distintas representaciones de los agrupamientos. Las figuras mostradas a continuación corresponden a las regiones obtenidas para todos los valores de k escogidos (en este caso 2, 3 y 4).

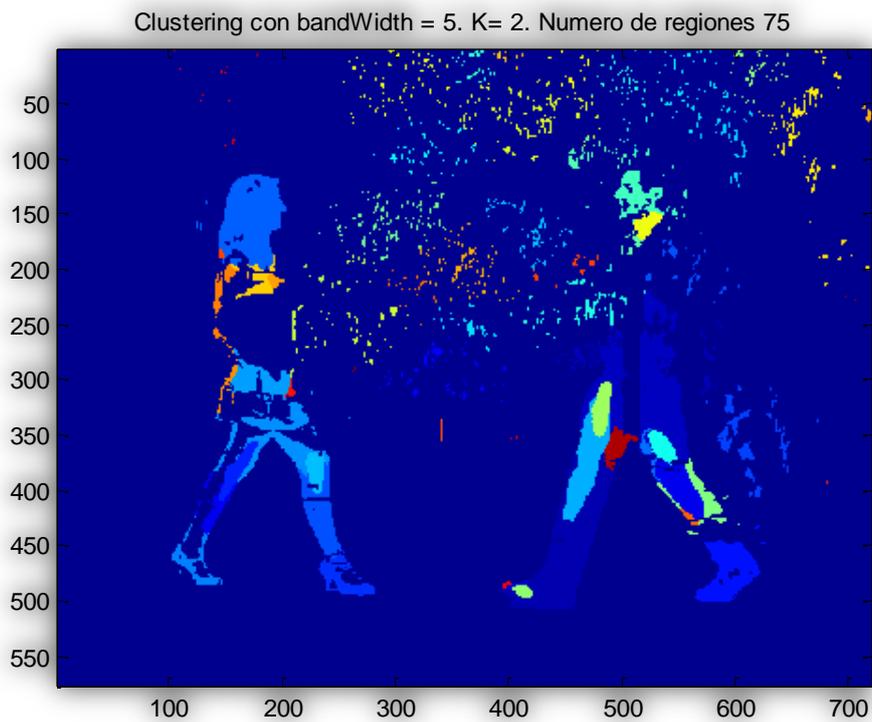


Figura 32: Clustering para $k = 2$

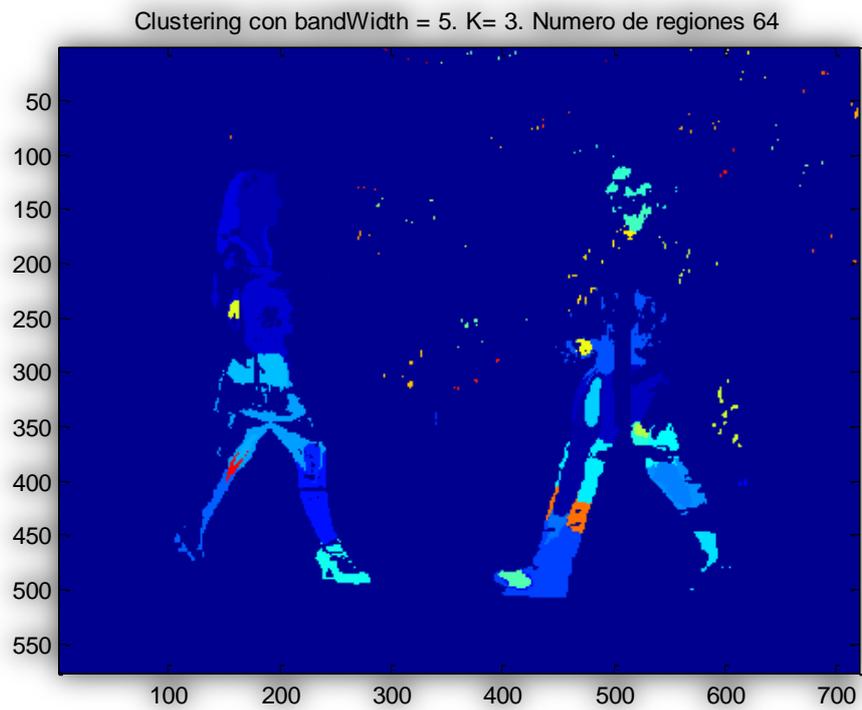


Figura 33: Clustering para k = 3

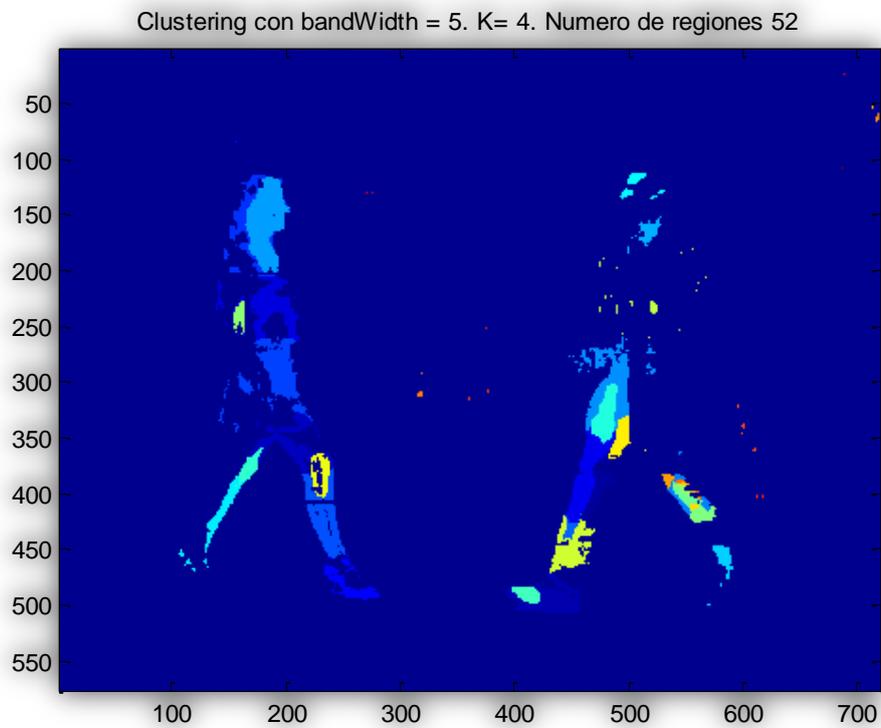


Figura 34: Clustering para k = 4

Por todo ello, se ha diseñado un método que sea capaz de discernir el valor de k más oportuno para cada una de las zonas de la imagen. El algoritmo propuesto realiza un análisis de cada una de las regiones que se han detectado mediante el primer proceso de *Mean Shift* con el valor más alto de k (el que dará lugar máscaras con menos píxeles) y realiza la identificación de las mismas regiones para los diferentes valores de k analizados siguiendo con un sistema de análisis y decisión que se explicará a continuación.

4.4.2. Subdivisión de regiones segmentadas

El algoritmo comienza recibiendo el primer grupo de regiones pertenecientes al *clustering* de los vectores de movimiento calculados sobre la máscara con valor de k mayor. Para cada una de las regiones determinadas en el primer proceso de *clustering* se realiza el siguiente procedimiento:

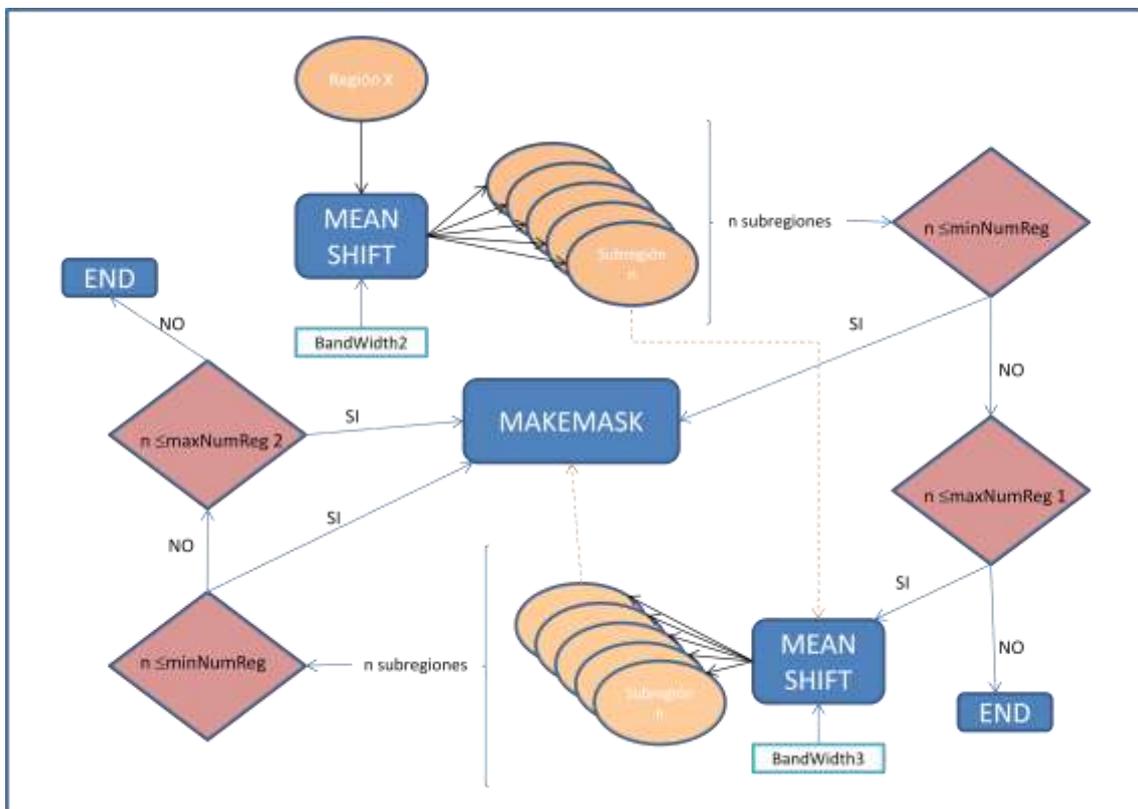


Figura 35: Esquema del primer proceso realizado para cada región.

Para cada una de las regiones se realiza un *clustering* que determinará el número y la naturaleza de las subregiones que la forman. Este segundo proceso de agrupamiento se realiza con un ancho de banda diferente (generalmente mayor) al utilizado en el primer proceso de agrupamiento (denominado **Bandwidth2**). Esto dará lugar a regiones menores cuya coherencia de movimiento ha sido demostrada en el primer agrupamiento. Lo que interesa es conocer el grado de coherencia que muestra la región al estudiarla independientemente de las otras regiones.

Mediante este segundo proceso de *Mean Shift* obtenemos un número determinado de subregiones. Este número debe ser comparado con los márgenes que estimamos oportunos para cada secuencia para determinar si el movimiento es coherente en la región. Si se considera que es coherente (número de regiones menor que el mínimo) se procede a la generación de la máscara final en la que se incluye la región estudiada y además se comienza el estudio de la región equivalente en la representación de las regiones de las máscaras con un valor de k inmediatamente inferior (proceso denominado en el esquema **MakeMask**). Si por el contrario se considera que no presenta el grado necesario de coherencia (número de regiones entre el mínimo y el máximo permitido), se realiza un tercer proceso de agrupamiento. Esta vez realizado sobre las subregiones de cada una de las regiones. Si la región no cumple ninguna de las premisas anteriores, y presenta un número de regiones mayor al máximo permitido, se procede al análisis de otra región, desechando la influencia de la misma en la máscara final.

Para el tercer proceso de agrupamiento se realiza el mismo procedimiento de selección, únicamente se modifican los valores máximos permitidos para cada subregión. Esto se debe a que los tamaños de las regiones estudiadas en este proceso son mucho menores a las iniciales, y por tanto deben obtener una coherencia mucho mayor para ser consideradas regiones de interés. Tras este proceso de selección se procede a la etapa de estudio de las regiones equivalentes en máscaras con k inmediatamente inferior. Si el resultado de la comparación es negativo se desechan las regiones.

Para ilustrar el procedimiento descrito, se toma por ejemplo la región 30 (del cuadro 140 de la secuencia Baile Multimodal) obtenida tras el primer proceso de *clustering* con un valor de ancho de banda (BandWidth1) igual a 5. Esta región entraría en el primer proceso explicado en la siguiente figura:

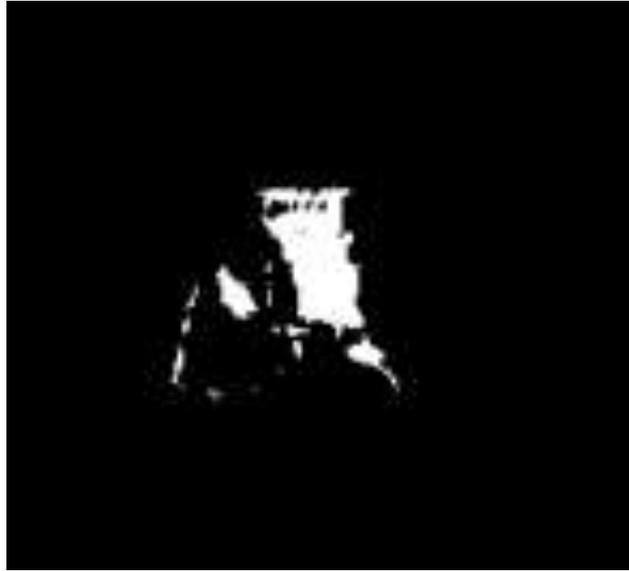


Figura 36: Representación de la región 30

Sobre la región escogida se realiza un segundo proceso de *clustering*, esta vez utilizando un valor de ancho de banda (BandWidth2) igual a 8 (valor óptimo para la secuencia BAILE). En la imagen siguiente se puede observar como la región presenta tres subregiones de distinto tamaño. El primer paso sería determinar si el número de subregiones está dentro de los márgenes óptimos para la secuencia a estudio. Estos valores se analizarán en profundidad para cada secuencia en el tema realizado para la asignación de los parámetros óptimos de cada una de las secuencias de video estudiadas en el proyecto.

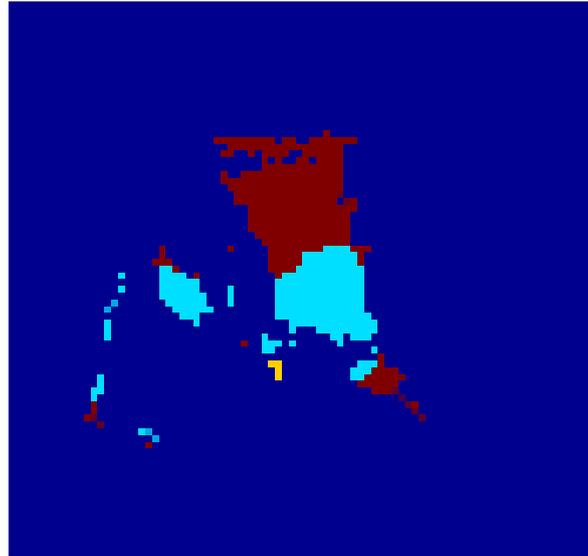


Figura 37: Representación de las subregiones de la región 30.

En este caso el número de subregiones es igual a tres. Lo cual indicará que la región puede ser considerada coherente si el número de subregiones es menor o igual a **minNumReg**, o deberá estudiarse en mayor profundidad si el número de subregiones está entre **minNumReg** y **maxNumReg1**. Suponiendo los siguientes valores:

$\text{minNumReg} = 2.$

$\text{maxNumReg1} = 8.$

$\text{maxNumReg2} = 3.$

En este caso, dado que el número de subregiones es 3, se deben analizar independientemente para comprobar su naturaleza. Al realizar el tercer proceso de agrupamiento de la región obtenemos los siguientes resultados para cada una de las 3 subregiones:

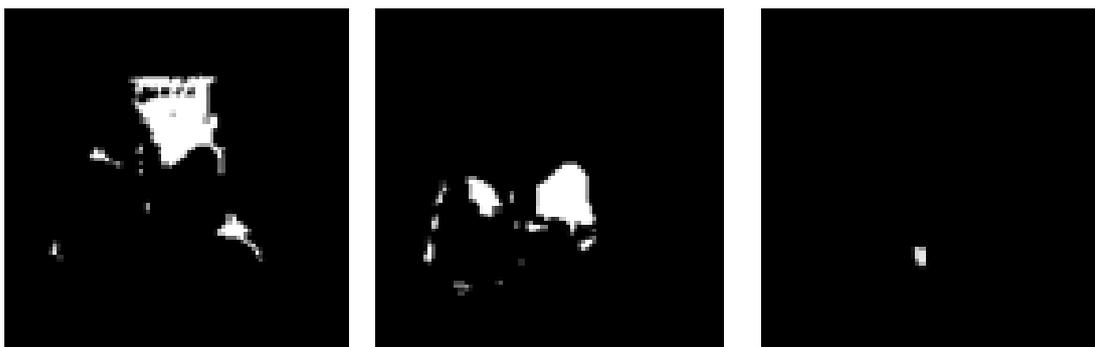


Figura 38: Representación de las subregiones de la región 30.

Para las regiones obtenidas se realiza el tercer y último proceso de Mean Shift del algoritmo, el cual determinarán las subregiones con un movimiento coherente entre sí. Se debe realizar el proceso de agrupamiento con un valor de ancho de banda mayor (denominado **BandWidth3**), acorde con las características de las regiones, mucho mas pequeñas que las estudiadas anteriormente,

Los resultados que se obtienen para las tres subregiones son mostrados en las siguientes figuras:

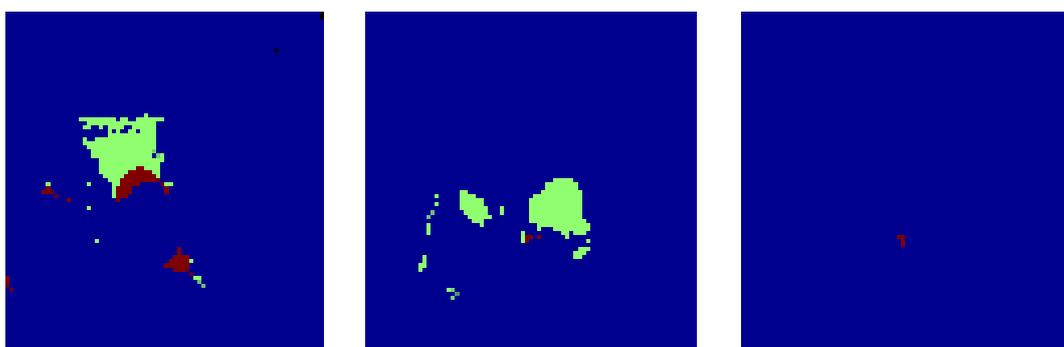


Figura 39: Representación de las subregiones de las subregiones obtenidas.

Para cada una de las subregiones estudiadas debemos comparar el número de las subregiones que a su vez las forman con el valor de la variable **maxNumReg2**. Si la comparación nos indica que el número de regiones es menor o igual al valor que se ha determinado (se realizarán múltiples pruebas y se obtendrá el resultado óptimo de los parámetros), concluiremos que la subregión forma parte de la máscara final

y se continuará con el estudio de la región equivalente para un valor de k inmediatamente menor.

4.4.3. Generación de la máscara de segmentación

El proceso de nominado **MakeMask** (que en nuestro código será denominado *MakeFinalMask*) consiste en dos subprocesos independientes. Por una parte, toda subregión que ha sido aceptada, es directamente **integrada en la máscara final** que se ha creado y que se irá completando con las regiones que se estimen coherentes en su movimiento. Este primer subproceso puede integrar directamente una región completa resultado del primer proceso de *clustering* o solo una de las muchas subregiones que pueden resultar tras la realización del segundo o tercer proceso de *clustering*. Por otra parte, el proceso **MakeMask**, desemboca en el **siguiente proceso de selección**, esta vez mediante el estudio de la región o subregión aceptada, que **es identificada** en la representación de regiones que se ha obtenido con un valor de k inferior. Es decir, toda región o subregión que es aceptada para el estudio (ha cumplido las especificaciones marcadas por el número máximo y mínimo de subregiones), será identificada en la representación que se tiene del estudio de Mean Shift para un valor de k inferior.

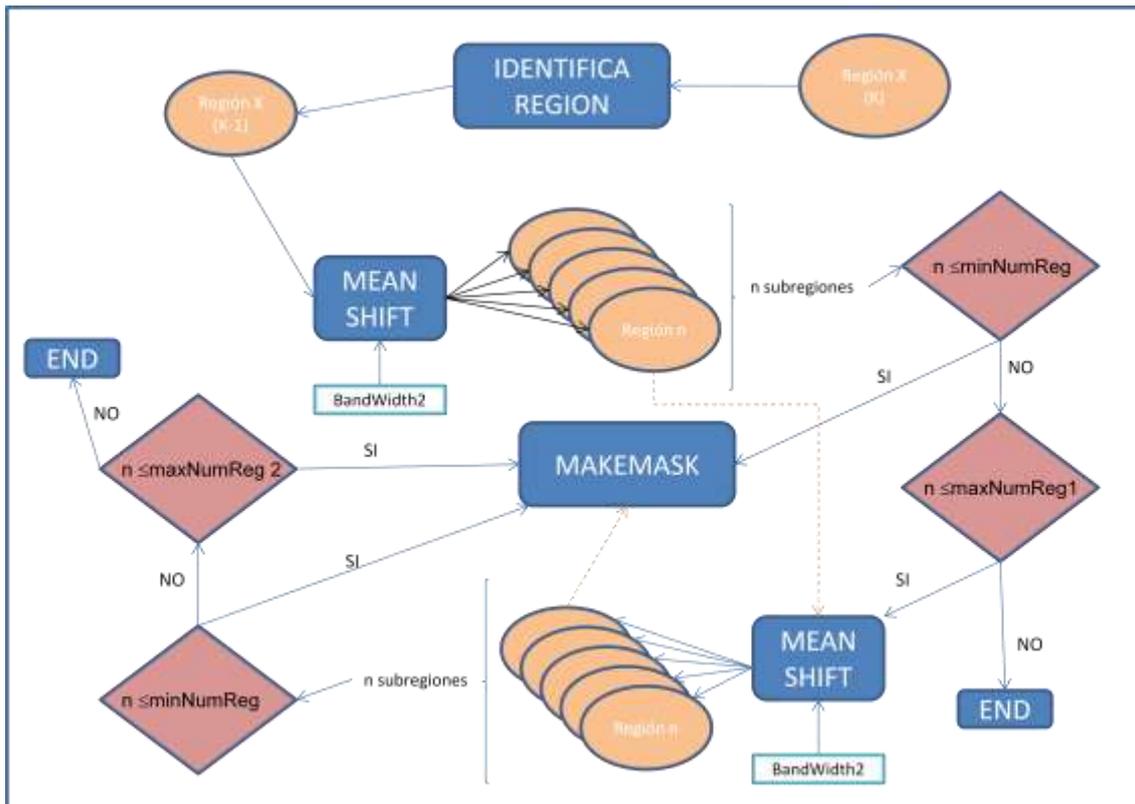


Figura 40: Diagrama de procesos realizados en MakeMask

Este proceso es equivalente al realizado para las regiones obtenidas para la representación con la k más alta, y se repetirá tantas veces como valores distintos del k inicial se hayan considerado. Es decir, si tenemos como valores de k 2, 3 y 4, el proceso de *MakeMask* se repetirá dos veces.

En este proceso se realiza el mismo análisis que hemos explicado anteriormente, con la diferencia de que las regiones a estudio son las regiones identificadas como equivalentes (es decir, las que ocupan la misma posición) en las representaciones para distintos valores de k .

La **identificación** se realiza mediante la búsqueda de todos los puntos de una región en la representación con el valor de k menor. Como cada punto de la matriz de representación tiene el valor de la región a la que pertenece, se buscan todas las regiones de la representación equivalente que tienen en su interior alguno de los puntos pertenecientes a la región que estamos estudiando. Como resultado, normalmente se obtiene una región mayor en tamaño (debido a que un valor

menor de k establece mayores regiones de movimiento) o un conjunto formado por varias regiones. Éstas serán estudiadas a su vez con el método descrito.

Para realizar una explicación más gráfica, tomamos por ejemplo la región 30 de la imagen obtenida de la secuencia BAILE. Esta región ha sido analizada y se tienen las 3 subregiones consideradas coherentes. A continuación se debe comparar el número de subregiones obtenidas con los valores de decisión que han sido definidos con anterioridad. Este proceso trae consigo la identificación de las regiones equivalentes en las máscaras con distinto k . El proceso de identificación de cada una de las secuencias daría como resultado una región, o un conjunto de regiones, situadas en las mismas zonas de la imagen, pero en las máscaras con un valor de k menor. Se obtienen las siguientes regiones:

Para el caso de la primera subregión extraída anteriormente se deben identificar las regiones equivalentes en la representación de las regiones para $K=3$. En este caso, se obtienen las siguientes regiones:

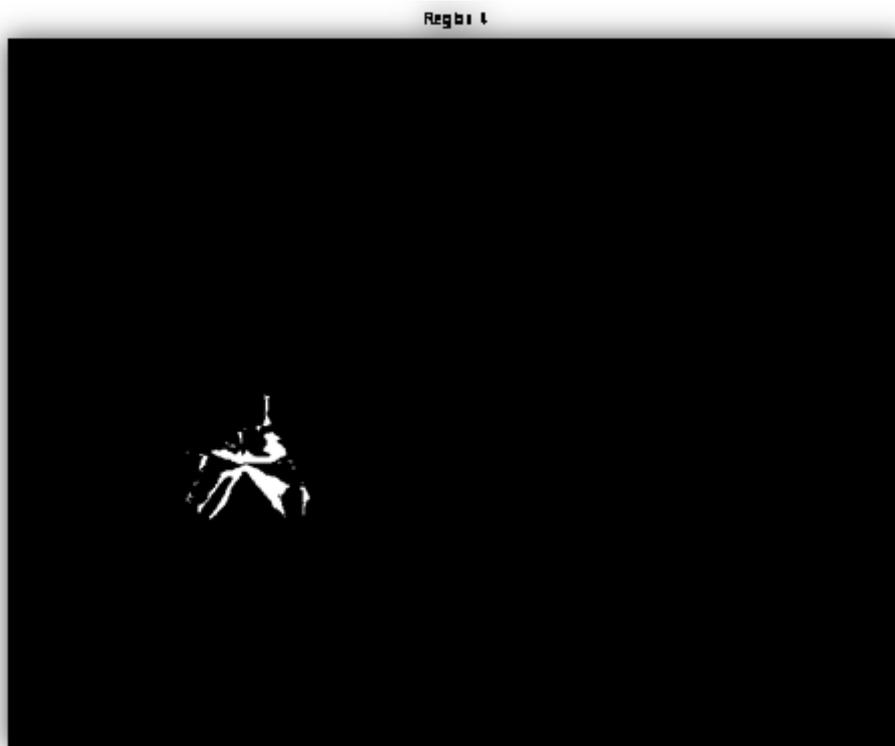


Figura 41: Región 4 identificada como equivalente en la representación con $K=3$.

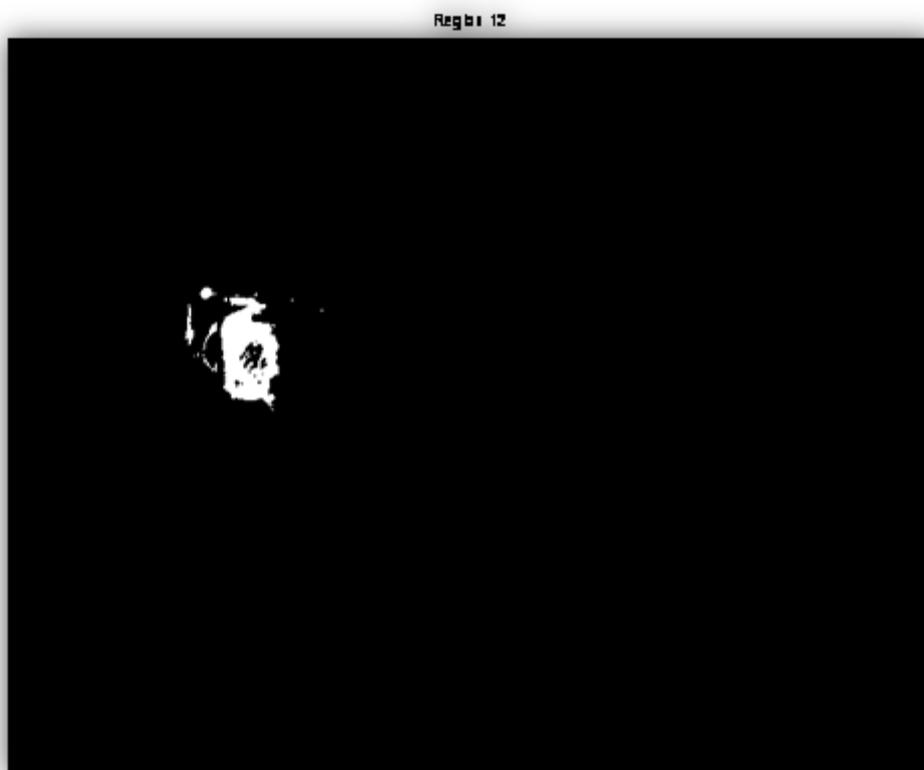


Figura 42: Región 12 identificada como equivalente en la representación con $K=3$.

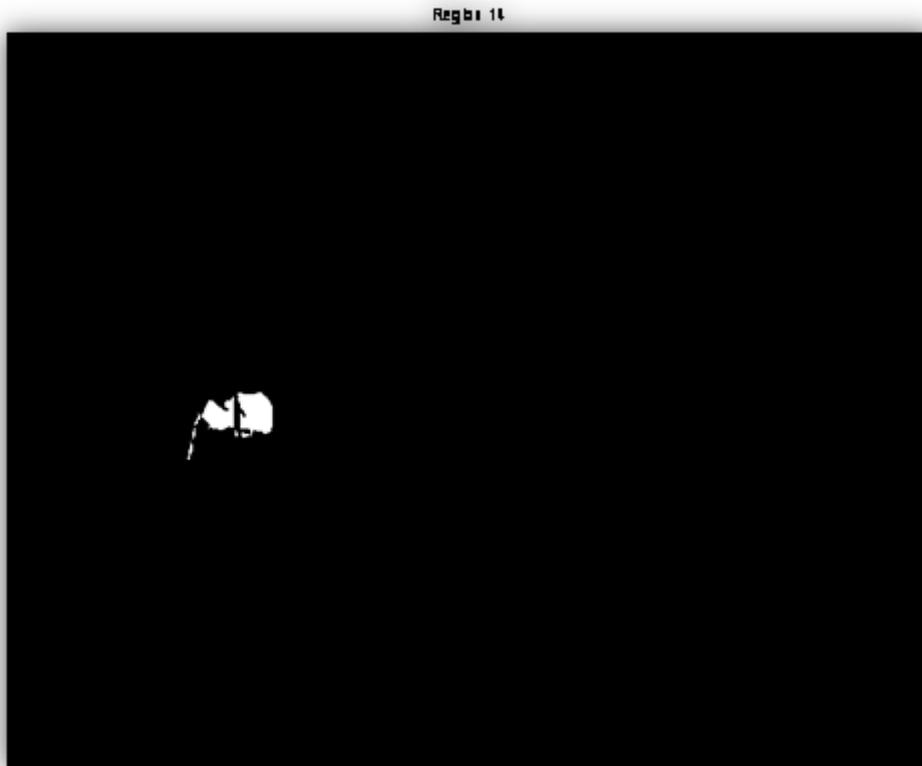


Figura 43: : Región 14 identificada como equivalente en la representación con $K=3$.

El resultado son las regiones 4, 12, y 14 de la representación de las regiones de movimiento para $K = 3$. Cada una de las regiones que han sido identificadas sería analizada de la misma manera que las regiones obtenidas para el primer proceso de agrupamiento. Se analizarán las subregiones que las forman y las que obtengan un resultado positivo en la comparación formarán parte de la máscara final y volverán a ser identificadas en la máscara obtenida con el valor de k inmediatamente menor, hasta completar el algoritmo completo, analizando todas las máscaras creadas con los distintos valores de k .

En las figuras siguientes se muestra cómo se realiza el proceso de generación de la máscara final. Como se ha visto a lo largo del capítulo, para cada una de las regiones obtenidas inicialmente se realiza un estudio exhaustivo de la naturaleza de los píxeles que la forman. En el caso de que las comparaciones hayan sido exitosas, la región o subregión estudiada se añade a una máscara final creada para conformar la máscara final. En las siguientes dos figuras, se muestra el resultado de la máscara en el momento de la finalización del estudio para la región número 6,

y posteriormente, tras la evaluación de la región 13. Es decir, en la primera imagen aparecen los resultados de la evaluación de las 6 primeras regiones, y en la segunda el resultado de las 13 primeras regiones estudiadas.



Figura 44: Resultado de segmentación de movimiento iteración 6.



Figura 45: Resultado de segmentación de movimiento iteración 13.

5

PRUEBAS Y RESULTADOS

En el presente capítulo se evaluarán los métodos que se han implementado a lo largo del proyecto. Para analizar los métodos estudiados, existen diversas estrategias de evaluación de la calidad de los sistemas de segmentación. Existen dos puntos de vista para enfocar el estudio de la calidad del sistema: los métodos subjetivos (54), en los cuales se depende de ciertos criterios humanos de evaluación, y métodos objetivos (55). Estas últimas técnicas utilizan una segmentación de referencia, que es la segmentación *ground truth*, la segmentación ideal, con el fin de compararla con la segmentación alcanzada con los procesos no ideales. La comparación se puede utilizar para comparar el número de píxeles erróneo y correctos detectados por el sistema de segmentación propuesto, así como para obtener otras medidas de calidad de la evaluación.

Las técnicas de evaluación utilizadas en este proyecto son técnicas objetivas, que se basan en una segmentación de referencia calculada para el conjunto de secuencias que se analizarán y que constituyen el grupo de videos de prueba ('*dataset*') utilizado en el proyecto.

5.1. Descripción del dataset utilizado

El conjunto de secuencias de videos de prueba que se ha utilizado para la evaluación de los algoritmos de segmentación han sido aportados por el *Video Processing and Understanding Lab* de la Escuela Politécnica Superior de la UAM (53) y están disponibles en la página siguiente web: <http://www-vpu.ii.uam.es/CVSG/>.

Para realizar las pruebas, se deben utilizar una serie de secuencias específicas, clasificadas por su complejidad o grado de dificultad que presenta para un algoritmo de segmentación su análisis. Esta dificultad depende de una serie de características de las secuencias que se denominan ‘factores críticos’ entre las que se encuentran las características específicas de los objetos, el tipo de fondo, o el movimiento de los mismos, vistas en profundidad en el Proyecto Fin de Carrera realizado por Sonsoles Herrero (13).

Los videos de prueba utilizados para evaluar los algoritmos de segmentación implementados en este proyecto, se han clasificado en base a las características descritas anteriormente y que influyen en los algoritmos de sustracción de fondo implementados. Además, se utilizará un identificador (S1, S2, ...) para agrupar a cada conjunto de secuencias de características similares (con el mismo tipo de fondo). Los videos utilizados para evaluar los métodos de segmentación desarrollados son los siguientes:

Secuencia	Identificador	Situación (interior/exterior)	Modalidad del fondo
 BAILE UNIMODAL	S1	Interior	Unimodal
 BAILE MULTIMODAL	S2	Exterior	Multimodal (alto)
 TENIS	S3	Exterior	Multimodal (medio)
 INGRAVIDEZ	S4	Exterior	Multimodal (alto)
 SILLA	S5	Exterior	Multimodal (alto)

Tabla 1: Secuencias seleccionadas para el estudio

5.2. Diseño de pruebas experimentales

Para obtener los parámetros de funcionamiento más adecuados de cada uno de los algoritmos implementados, se han estudiado y ejecutado una serie de pruebas experimentales. Estas pruebas consisten en la evaluación de todos los métodos utilizados para una serie de parámetros que resultan fundamentales para la optimización de los resultados.

Para el método de segmentación las pruebas consisten en variaciones de los parámetros de los que depende el algoritmo suponiéndoles independientes para cada una de las secuencias analizadas, con el fin de fijar unos valores coherentes para cada una de ellas dependiendo de sus características.

Los rangos de prueba se han escogido en función de los factores críticos de las secuencias; por ejemplo, las secuencias de fondo multimodal exigen mayor variación en el fondo y, por tanto, varianzas (σ^2 , la desviación elevada al cuadrado) más elevadas que aquellas secuencias cuyo fondo sólo tenga un tipo de movimiento. Por otro lado, las secuencias con objetos en movimiento lento requieren un factor de aprendizaje (α) más bajo que aquellas en los que los movimientos de los objetos sean rápidos. Aplicando los rangos de prueba a cada parámetro, se han ido encontrando los parámetros más adecuados de cada método de modelado de fondo.

A continuación, se muestra una tabla que recoge los parámetros más representativos y los rangos de prueba con los que se han realizado las ejecuciones sobre las secuencias del apartado anterior. En el caso de la segmentación basada en la Gaussiana Simple, los dos campos de parámetros introducidos para realizar las pruebas son los siguientes:

- El primer vector hace referencia a los valores utilizados para establecer el valor más preciso de la **desviación**, siendo en nuestro caso valores comprendidos entre 1 y 50, y han sido introducidos como parámetros en

las pruebas que se ha realizado para el proceso de prueba preliminar de parámetros. Se debe indicar que los valores que se obtuvieron en el Proyecto Fin de Carrera en el que se basan estas pruebas (13), han sido modificados para ajustarlos mejor a los resultados que se han obtenido en este proyecto.

- El segundo vector pertenece a los valores de la **alpha** que se han introducido en las pruebas. Estos valores varían desde 10^{-8} hasta valores de 10^{-1} . Las pruebas realizadas en este caso dependen de si la secuencia es unimodal o multimodal.

Algoritmo	Valores de la desviación	Valores de alpha
Gaussiana Simple	[1:2:18]	$[10^{-4}, 5 \cdot 10^{-4}, 10^{-3}, 5 \cdot 10^{-3}, 10^{-2}, 5 \cdot 10^{-2}, 10^{-1}]$
	[20:5:40]	$[10^{-8}, 5 \cdot 10^{-7}, 10^{-7}, 5 \cdot 10^{-6}, 10^{-6}, 5 \cdot 10^{-5}, 10^{-5}]$

Tabla 2: Parámetros de prueba para la segmentación

Posteriormente se han creado y ejecutado una serie de pruebas para comparar los métodos de estimación de movimiento implementados a lo largo del proyecto. Se debe tener en cuenta que con realizar un estudio comparativo se pretende observar las características reales que ofrecen los métodos en las secuencias de imágenes seleccionadas. Para ello, se **probarán los métodos de flujo óptico y de Block Matching** para determinar sus parámetros ideales y **se seleccionará el método o métodos que mejores resultados ofrezcan.**

ALGORITMO BASADO EN FLUJO ÓPTICO			
Algoritmo	Número de niveles	Alpha global	Alpha local
Flujo óptico	$(\log_{10}(30/\min(h,w))/\log_{10}(0.9))$, w=ancho de la imagen h = altura de la imagen	[8:2:14]	[10:5:15]

Tabla 3: Parámetros de prueba para la estimación basada en Optical Flow

Primeramente se establecerán los intervalos de análisis para los valores de los parámetros de los algoritmos Block Matching, excluyendo el método *Four Step Search*, el cual ya establece en su ejecución los parámetros óptimos.

ALGORITMOS BLOCK MATCHING		
Algoritmo	Tamaño de bloque	Máximo desplazamiento
Exhaustive Block Matching	[4,8,16]	[4,4,16]
Three Step Search	[4,8,16]	[4,8,16]
New Three Step Search	[4,8,16]	[4,8,16]
Simple and Efficient TSS	[4,8,16]	[4,8,16]
Diamond Search Method	[4,8,16]	[4,8,16]

Tabla 4: Parámetros de prueba para la estimación basada en desplazamiento de bloques

Para el caso de los algoritmos de agrupamiento implementados, también se debe tener en cuenta la influencia de los parámetros en los resultados obtenidos. Para

ello, se debe afinar la búsqueda de los parámetros óptimos mediante pruebas realizadas en las secuencias existentes en nuestra base de datos.

Parámetros	ALGORITMO MEAN SHIFT
Relación: Valor de los vectores / Area	[1:0.1:2]

Tabla 5: Relación entre parámetros de Mean Shift

Parámetros	ALGORITMO MEAN SHIFT
Bandwidth1	[5:1:8]
Bandwidth2	[6:1:10]
Bandwidth3	[6:1:10]

Tabla 6: Rango de parámetros de Mean Shift para la agrupación de los vectores de movimiento

En el proceso recursivo que identifica y decide si las regiones mantienen un movimiento coherente en todas las subregiones, se debe limitar un número máximo o mínimo de subregiones en las cuales el movimiento puede resultar ligeramente diferente (se dice ligeramente diferente debido a que para formar parte de una región inicial, los vectores de las regiones ya deben poseer un parecido importante). Para ello, se han definido unos valores que determinan para cada región el valor máximo y mínimo en cada uno de los procesos de Mean Shift utilizados. Los parámetros son: **minNumReg**, **maxNumReg1** y **maxNumReg2**, como se ha visto en el apartado correspondiente a la explicación del método utilizado.

Debido a la diferente naturaleza de las secuencias analizadas, se han realizado experimentos para cada una de ellas, en los que se determinan los valores mediante los cuales se han obtenido unos resultados más precisos. Los valores utilizados en las pruebas han sido los siguientes:

Parámetro	Valores de prueba
minNumReg	[2:1:4]
maxNumReg1	[6:1:9]
maxNumReg2	[3:1:6]

Tabla 7: Rango de parámetros de prueba para la segmentación de movimiento.

5.3. Métricas utilizadas

Las métricas que se han utilizado para evaluar los métodos de segmentación desarrollados en este PFC son medidas objetivas que se basan en la comparación de las máscaras binarias del frente generadas por cada método y el 'ground-truth'. Aunque se han encontrado distintas medidas perceptuales en la literatura [82] finalmente se ha optado por las medidas objetivas a nivel de píxel que se recuerdan 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 del 'grund truth' y la máscara obtenida a la salida del algoritmo de segmentación.

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

Tabla 8: Parámetros estadísticos de la evaluación

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 detectados por nuestro algoritmo de segmentación de ese tipo. Las máscaras que se obtienen de la segmentación son binarias presentando píxeles con valor cero y valor uno. Por otro lado, nuestro objetivo a la hora de evaluar los algoritmos no es únicamente la correcta detección del objeto sino también, que los píxeles en ausencia de movimiento sean correctamente detectados, esto supone el cálculo de precisión en los dos grupos:

$$\text{Precisión (píxeles de valor 0)} = P0 = \frac{TN}{TN + FN}$$

$$\text{Precisión (píxeles de valor 1)} = P1 = \frac{TP}{TP + FP}$$

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

$$\text{Recall (píxeles de valor 0)} = R0 = \frac{TN}{TN + FP}$$

$$\text{Recall (píxeles de valor 1)} = R1 = \frac{TP}{TP + FN}$$

- **Fscore:** 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. El f-score también ha sido calculado para cada conjunto de valores de la máscara binaria resultado del segmentador:

$$Fscore(\text{píxeles de valor } 0) = FS0 = \frac{2 \cdot P0 \cdot R0}{P0 + R0}$$

$$Fscore(\text{píxeles de valor } 1) = FS1 = \frac{2 \cdot P1 \cdot R1}{P1 + R1}$$

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 (unos y ceros) fusionando las medidas anteriores a través de una suma (si bien se podía haber utilizado cualquier otra función). La selección del parámetro óptimo se realiza eligiendo la suma máxima:

$$SUMA = FS0 + FS1$$

5.4. Selección de método óptimo de estimación de movimiento

Para la estimación de movimiento, se debe realizar un estudio de los seis métodos de estimación de movimiento basados en el desplazamiento de bloques y del método de flujo óptico implementado. De esta manera, gracias al conocimiento que nos ha aportado su estudio, se obtiene el mejor método de los presentes en el estado del arte para realizar el cometido del proyecto.

Se deben realizar comparativas referentes a la calidad de los resultados y al tiempo de ejecución mediante las cuales se pueda discernir cual o cuales de los métodos presenta resultados considerablemente mejores a los demás.

A continuación, igual que para el método de segmentación, se han realizado las pruebas oportunas con el fin de establecer cuáles de los métodos consiguen mejores resultados. Las pruebas han sido realizadas mediante la variación de los parámetros establecidos en el apartado 5.2, concretamente en la Tabla 3 y en la Tabla 4, y han sido evaluadas en el momento de la implementación observando cuáles de los parámetros resultan más oportunos para cada uno de los métodos de

estimación de movimiento. Esta evaluación se ha realizado de dicha manera teniendo en cuenta que los resultados de la estimación de movimiento son resultados apreciables y comparables visualmente. Por lo tanto en el proceso de implementación de los algoritmos de estimación, se ha realizado una evaluación para una de las secuencias a estudio, la secuencia TENIS. Se ha escogido esta secuencia debido a las características del movimiento de la misma, dado que al ser una secuencia multimodal media tiene validez en el estudio de las secuencias tanto multimodales altas como unimodales.

5.4.1. Tiempos de ejecución

En primer lugar se ha realizado una comparativa en coste computacional (tiempo de ejecución) de los algoritmos implementados, para dos fotogramas determinados de una de las secuencias seleccionadas en el estudio, la secuencia TENIS.

Los tiempos se han expresado en segundos por cada una de las estimaciones de movimiento entre dos imágenes consecutivas de la secuencia. Las imágenes estudiadas son imágenes de tamaño 720x576 píxeles, en formato '.png'.

ALGORITMO BASADO EN FLUJO ÓPTICO			
Algoritmo	Número de niveles	Alpha global	Alpha local
Flujo óptico	$(\log_{10}(30/\min(h,w))/\log_{10}(0.9))$, w = ancho de la imagen = 720. h = altura de la imagen = 576.	10	15

En la siguiente tabla quedan excluidos los parámetros del algoritmo *Four Step Search*, como se ha explicado anteriormente.

ALGORITMOS BLOCK MATCHING		
Algoritmo	Tamaño de bloque	Máximo desplazamiento
Exhaustive BM	4	4
Three Step Search	4	8
New Three Step Search	4	16
Simple and Efficient TSS	4	16
Diamond Search Method	4	8

Tabla 9: Parámetros óptimos para la estimación de movimiento basada en BM

Con los parámetros anteriores, la estimación de movimiento basada en cada uno de los métodos presentaba las siguientes duraciones:

Tiempo de procesamiento para cada imagen (segundos)							
Secuencia	Flujo óptico	EBM	TSS	NTSS	SETSS	FSS	DS
S1 Fotogramas 209-210	214.2	218.234	68.469	61.813	36.468	72.454	100.187
S2 Fotogramas 199-200	212.36	209.562	64.609	58.484	36.454	70.656	100.812
S3 Fotogramas 139-140	210.25	210.438	65.703	59.641	36.453	72.765	105.375
S4 Fotogramas 109-110	209.36	210.625	65.250	59.266	37.50	70.672	99.640

S5							
Fotogramas 39-40	207.21	208.563	64.547	58.937	36.078	70.031	100.344

Tabla 10: Tiempo de procesamiento de los algoritmos de estimación de movimiento

Se observa como los tiempos de ejecución de los algoritmos son altos, debido a la cantidad de píxeles que conforman la imagen y al número de operaciones que necesitamos para obtener estimaciones precisas para los parámetros que hemos considerado óptimos en cada método.

También se estudia el número de operaciones que se realiza en cada una de las iteraciones del algoritmo, que da una idea de la carga computacional que presenta cada uno de los algoritmos. Esta medida únicamente se puede realizar a los algoritmos de estimación de movimiento basados en el desplazamiento de bloques, debido a que se quiere calcular el número de operaciones por bloque y el método del flujo óptico no presenta este modo de cálculo.

Número medio de puntos analizados para cada bloque						
Secuencia	EBM	TSS	NTSS	SETSS	FSS	DS
S1						
Fotogramas 209-210	80.1025	24.7316	22.5303	12.5391	26.1600	35.4261
S2						
Fotogramas 199-200	80.1025	24.7136	22.2225	12.2400	26.4920	36.6730
S3						
Fotogramas 139-140	80.1025	24.7135	22.0560	12.1980	26.5432	36.8431

S4 Fotogramas 109-110	80.1025	24.7147	22.2199	12.4503	26.2667	35.9096
S5 Fotogramas 39-40	80.1025	24.7159	22.2110	12.2690	26.3934	36.4204

Tabla 11: Número de puntos analizados para cada bloque

En los datos arrojados se observa como para el caso del método exhaustivo, el número de operaciones resulta siempre el mismo, dado que siempre se realizan las operaciones máximas posibles, y de entre los ellas, se selecciona el mejor resultado. Se observa también como el método que menos carga computacional ofrece es el método Simple y Eficiente basado en Three Step Search.

5.4.2. Calidad de la estimación de movimiento

En el presente Proyecto Fin de Carrera se ha ido desarrollando una serie de métodos de estimación de movimiento, con el fin de comprobar cuál de ellos ofrecía mejores resultados. Con las pruebas que se han ido realizando (véase apartado 5.2), se ha creído conveniente centrarnos en el método que ofrece unos resultados más precisos, aun teniendo como hándicap el tiempo de ejecución del mismo. De esta manera, los resultados que se han ido obteniendo al ejecutar los algoritmos de estimación mostraban que el método basado en el flujo óptico resultaría considerablemente más preciso a la hora de influir positivamente en las decisiones que se llevara a cabo en el proceso de segmentación basado en la información del modelado de fondo.

5.4.3. Conclusiones

Con las pruebas que se han realizado se llega a la conclusión que el método que mejor compensa la velocidad de ejecución o de cálculo con los resultados más precisos es el basado en el flujo óptico. Con este método, se obtiene una precisión a nivel subpíxel, que nos ofrece una visión completamente fiable del movimiento que experimenta un objeto en la imagen, calculando los vectores hasta en situaciones para las cuales los métodos basados en el desplazamiento de bloques no pueden estipular el movimiento.

Otro punto fundamental es que gracias al método de flujo óptico implementado, se pueden calcular los vectores de movimiento de las zonas homogéneas de la imagen, como pueden ser texturas de la ropa que presentan la misma tonalidad, en las que el cálculo de los vectores basándonos en desplazamiento de bloques resulta muy complicado.

Tras haber realizado un estudio de los métodos de estimación encontrados en el estado del arte, **se cree conveniente para la realización del método de integración de la información de movimiento con la segmentación, el uso del método basado en el flujo óptico**, debido que es el que más información nos proporciona, siendo fundamental un estudio preciso del movimiento de los objetos para evaluar posteriormente las características del mismo.

Pese que a priori se planteaba la ejecución del algoritmo propuesto para la integración con todos los métodos de estimación, la experiencia que ha aportado la realización de pruebas preliminares demuestra que al realizarlos con el método de flujo óptico se obtienen siempre mejores resultados que con los métodos de estimación basados en el desplazamiento de bloques.

Por todo ello, los procesos realizados a continuación han sido realizados con la estimación del movimiento basada en el flujo óptico.

5.5. Selección de parámetros

Tras haber realizado las pruebas preliminares enunciadas en el apartado 5.2, se obtienen una serie de parámetros independientes cuyos valores son los más óptimos para cada una de las secuencias que han sido estudiadas. Para cada una de ellas, se obtienen unos parámetros de segmentación, de estimación y de clustering diferentes, debido a las distintas propiedades de las secuencias. Los parámetros son los siguientes:

5.5.1. Parámetros para la segmentación

Tal y como se ha definido en el apartado 5.2, se han realizado una serie de pruebas preliminares para obtener para cada uno de los parámetros establecidos en la Tabla 2 unos valores que arrojen resultados precisos en la segmentación. Los resultados obtenidos para cada uno de los valores de los parámetros mostraban, mediante la máscara obtenida, la calidad de los mismos. Estas máscaras han sido consideradas y evaluadas de una manera sistemática y visual, observando cuales de los parámetros resultaban más eficientes a la hora de la realización de la separación entre frente y fondo, así como comparadas con los valores obtenidos en el Proyecto Fin de Carrera de Sonsoles Herrero (13).

El objetivo es determinar que parámetros presentan una sensibilidad mayor, es decir, aquellos en los que una pequeña variación altera en gran medida los resultados finales. Los resultados que mejor se adaptan a las secuencias escogidas han sido los siguientes:

Secuencia	Rango de Alpha	Alpha óptimo	Rango de desviación	Desviación óptima
S1	$[10^{-8}, 5 \cdot 10^{-7}, 10^{-7}, 5 \cdot 10^{-6}, 10^{-6}, 5 \cdot 10^{-5}, 10^{-5}]$	10^{-5}	[1:2:18]	20
S2	$[10^{-4}, 5 \cdot 10^{-4}, 10^{-3}, 5 \cdot 10^{-3}, 10^{-2}, 5 \cdot 10^{-2}, 10^{-1}]$	10^{-2}	[20:5:40]	25
S3	$[10^{-4}, 5 \cdot 10^{-4}, 10^{-3}, 5 \cdot 10^{-3}, 10^{-2}, 5 \cdot 10^{-2}, 10^{-1}]$	10^{-2}	[20:5:40]	10
S4	$[10^{-4}, 5 \cdot 10^{-4}, 10^{-3}, 5 \cdot 10^{-3}, 10^{-2}, 5 \cdot 10^{-2}, 10^{-1}]$	10^{-2}	[1:2:18]	10
S5	$[10^{-4}, 5 \cdot 10^{-4}, 10^{-3}, 5 \cdot 10^{-3}, 10^{-2}, 5 \cdot 10^{-2}, 10^{-1}]$	$5 \cdot 10^{-3}$	[1:2:18]	10

Tabla 12: Parámetros óptimos para las segmentación.

5.5.2. Parámetros de la estimación de movimiento

Los parámetros seleccionados para la estimación de movimiento han sido los seleccionados en el punto 5.4 para el método de estimación de movimiento basado en el flujo óptico.

5.5.3. Parámetros óptimos para el clustering

Para el desarrollo del algoritmo de segmentación de movimiento, se han utilizado distintos anchos de banda en el proceso de *clustering*. Esto se debe a que en cada una de las secuencias, el movimiento de los objetos o personas tiene una naturaleza distinta. En unas secuencias el movimiento de objetos externos al frente, como puede ser el movimiento de las hojas de los árboles, es mayor que el movimiento de las hojas de un arbusto que aparece en otra secuencia. Por ello, se debe otorgar un valor distinto a los anchos de banda del proceso en cada una de las secuencias.

Para conseguir unos resultados lo más precisos posible, se ha realizado un estudio exhaustivo de la influencia de cada uno de los parámetros del proceso de *clustering* en los resultados finales. De esta manera, obtenemos para cada una de las secuencias los valores de los anchos de banda (BandWidth1, BandWidth2 y

BandWidth3) que arrojan unos resultados más precisos. Estos valores han sido determinados en los intervalos propuestos en la Tabla 5 y 6 del apartado 5.2. Los resultados de todas las pruebas realizadas son mostrados en el Anexo 1, y en la siguiente tabla se observan los valores que mejores resultados han obtenido. El criterio de decisión ha sido la variable SUMA, la cual indica el grado de calidad del algoritmo en cuestión.

Parámetros	ALGORITMO MEAN SHIFT
Relación: Valor / Area	Valor de los vectores / Area = 1

Tabla 13: Relación entre parámetros de Mean Shift

Secuencia	Bandwidth 1	Bandwidth 2	Bandwidth 3
S1	6	8	8
S2	5	8	8
S3	5	7	6
S4	5	6	8
S5	5	6	6

Tabla 14: Parámetros óptimos para los procesos Mean Shift.

5.5.4. Parámetros óptimos para la segmentación de movimiento

Para la ejecución del algoritmo implementado, debemos fijar previamente los valores de decisión del mismo. Éstos, son los valores de los parámetros que determinan el grado de coherencia de una región, los denominados *minNumReg*, *maxNumReg1* y *MaxNumReg2* y han sido estudiados de manera exhaustiva, teniendo en cuenta los rangos establecidos en el apartado 5.2 (concretamente en la Tabla 7). Los resultados para todas las variables posibles se recogen en el Anexo 2.

Tras las pruebas realizadas se han asignado los siguientes valores. Estos parámetros son los utilizados para la evaluación de los resultados de cada una de las secuencias:

Secuencia	minNumReg	maxNumReg1	maxNumReg2
S1	2	8	5
S2	2	9	4
S3	1	9	6
S4	2	7	6
S5	2	8	5

Tabla 15: Parámetros óptimos para el proceso de segmentación de movimiento.

5.6. Análisis comparativo

5.6.1. Calidad de la segmentación

La mejora del método seleccionado de segmentación basado en la Gaussiana Simple tiene como objetivo eliminar píxeles o regiones en movimiento pertenecientes al fondo (como pueden ser árboles, arbustos, nubes, etc...). Esta integración de la información de movimiento en la segmentación inicial da como resultado una mejora considerable de las máscaras obtenidas, en las cuales desaparecen zonas que habían sido consideradas como fondo y aparecen zonas que no habían sido tenidas en cuenta y que realmente si pertenecen al frente de la imagen.

A continuación, se realizan las siguientes pruebas comparativas:

- Gráficas de porcentajes (expresados en tanto por uno) por cada imagen de cada secuencia de prueba de las siguientes medidas:

- Positivos incorrectos: son los píxeles de valor uno detectados incorrectamente, es decir, detección del objeto en movimiento (FG) incorrecta.
- Negativos incorrectos: son los píxeles de valor cero detectados incorrectamente, es decir, detección del fondo (BG) incorrecta.
- Tablas y diagramas de barras de las medidas de calidad (descritas en el apartado 5.3): precisión en cero (P0), precisión en uno (P1), recall en cero (R0), recall en uno (R1), fscore en cero (FS0) y fscore en uno (FS1). Todas estas medidas son las calculadas a todos los píxeles de cada imagen de las secuencias de prueba.

Los resultados de las pruebas representativas se muestran a continuación:

GAUSSIANA SIMPLE CON Y SIN MEJORA							
Secuencia	Tipo	P0	P1	R0	R1	FS0	FS1
S1	Sin mejora	0.939	0.991	0.999	0.569	0.968	0.723
	Con mejora	0.965	0.945	0.993	0.762	0.979	0.844
S2	Sin mejora	0.935	0.973	0.998	0.509	0.966	0.669
	Con mejora	0.953	0.964	0.997	0.653	0.974	0.778
S3	Sin mejora	0.942	0.889	0.997	0.253	0.969	0.394
	Con mejora	0.950	0.741	0.989	0.372	0.969	0.495
S4	Sin mejora	0.868	0.832	0.992	0.203	0.926	0.326
	Con mejora	0.888	0.907	0.993	0.336	0.938	0.491
S5	Sin mejora	0.943	0.849	0.988	0.536	0.965	0.657
	Con mejora	0.941	0.950	0.996	0.519	0.968	0.671

Tabla 16: Comparativa entre método con mejora y sin mejora

Los resultados mostrados en la gráfica anterior han sido seleccionados tras las múltiples pruebas que se han realizado a la hora de desarrollar el algoritmo. En el Anexo 2 se pueden observar los valores que se han obtenido para las distintas variables que han sido modificadas en cada uno de los procesos de evaluación.

5.6.1.1. Resultados por secuencias

Para una explicación más detallada del significado de los resultados mostrados en la Tabla 16, se va a analizar secuencia a secuencia las mejoras que se producen dependiendo de las características físicas de las mismas.

SECUENCIA 1: BAILE UNIMODAL

Los resultados más precisos que se han obtenido han sido los siguientes, para los métodos con y sin mejora:

	Sin mejora	Con mejora
TP	30771	41223
FN	23352	12900
TN	360313	358204
FP	284	2393
P0	0.939	0.965
P1	0.991	0.945
R0	0.999	0.993
R1	0.569	0.762
FS0	0.968	0.979
FS1	0.723	0.844
SUMA	1.691	1.823

Tabla 17: Resultados para la secuencia Baile Unimodal.

Para el caso de las secuencias unimodales, se pueden obtener resultados muy precisos mediante la utilización del método de segmentación basado en la *Gaussiana simple*, sin embargo el algoritmo que se propone en el proyecto aporta una mejora considerable en los resultados. Concretamente obtenemos una mejora importante en la detección de los objetos del frente. Con nuestro método, se

mejora en un 19,3% la detección de puntos de la imagen que según la máscara *ground truth* pertenecen al frente. Este valor se mide con el parámetro *Recall*, que indica la proporción de píxeles detectados como frente respecto al total de píxeles considerados como frente en la máscara ideal. Esto indica que casi un 20% de puntos del frente que no habían sido detectados se detectan a raíz del estudio que se ha realizado del movimiento que experimentan los píxeles de la imagen.

En la gráfica siguiente observamos la comparativa entre el método sin mejora y el método propuesto:

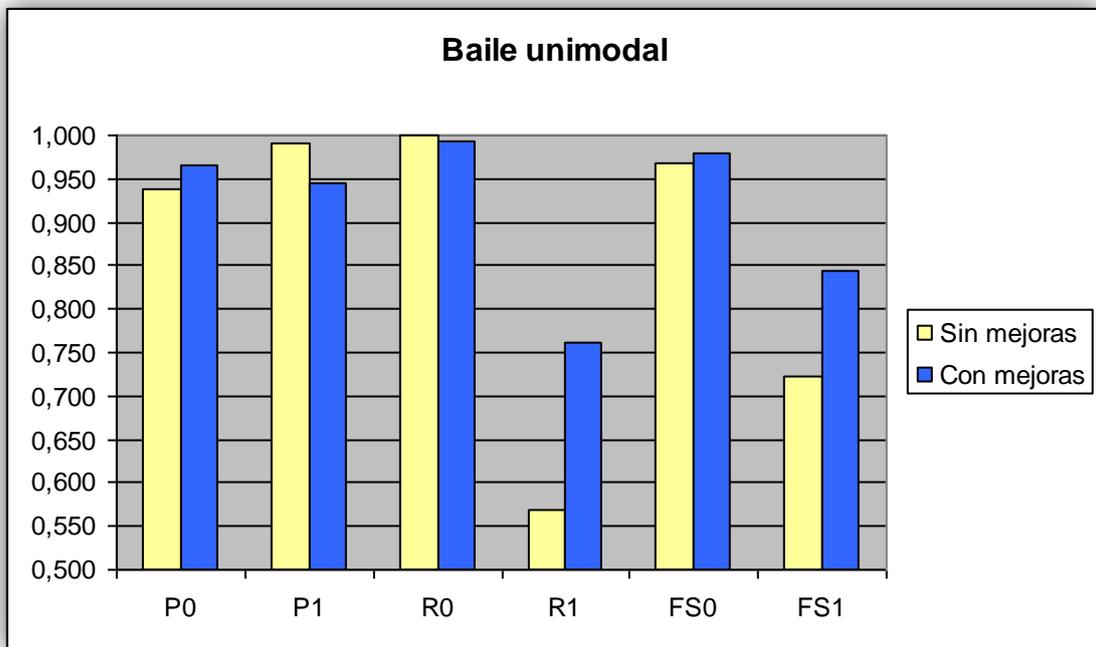


Figura 46: Diagrama de barras de los parámetros estudiados en la secuencia Baile Unimodal

En la gráfica se observa una mejora en el valor P0, que significa una mayor precisión de los valores detectados como fondo. El valor de R0 cercano a 1 (al 100% en medidas porcentuales) indica que la detección de los puntos del fondo es prácticamente perfecta. Se han detectado un 99,3% de los puntos del fondo.

Por otra parte, se observa que el valor de P1 disminuye con respecto al valor obtenido sin la mejora. Esto se debe a que en este caso se han detectado muchas más zonas como frente y se han obtenido zonas que con el algoritmo propuesto se

han considerado objetos del frente por sus características del movimiento pero que realmente no pertenecen al frente. Este error se compensa considerablemente si se comparan las zonas nuevas de frente detectadas con las zonas de frente establecidas por la máscara *ground truth*. La conclusión que se obtiene de estos datos es que se han detectado muchas más zonas de frente, y pese a que alguna de ellas no ha resultado precisa, se obtiene un 20 % más de zonas de frente correctas detectadas.

Los valores de *Fscore* calculados arrojan una medición que combina los resultados de la precisión y del *recall* calculados para ambos métodos. En el caso del método propuesto, se obtiene una mejora de la precisión tanto para la detección del frente como del fondo. Concretamente se obtiene un 1,2 % de mejoría para la detección del fondo y un 12,1 % de mejora en la precisión de la detección del frente.

A continuación se observa una comparativa cualitativa entre una primera máscara obtenida por el método basado en la *Gaussiana Simple* y una segunda, obtenida mediante el método propuesto.

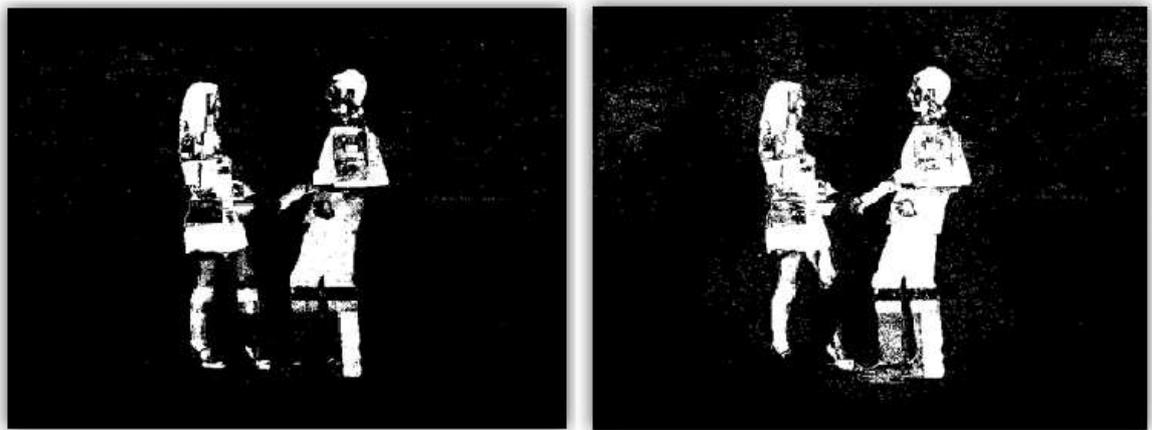


Figura 47: A la izquierda se muestra la máscara obtenida mediante el método basado en la *Gaussiana Simple* y a la derecha la obtenida con el método propuesto.

En la comparativa se observa como en la máscara obtenida con el método propuesto se obtiene mayor precisión en las zonas de las siluetas de las personas completando casi por completo su superficie. También observamos como aparecen

zonas del fondo que han sido consideradas como frente debido a sus condiciones de iluminación o de movimiento.

SECUENCIA 2: BAILE MULTIMODAL

Los resultados que se han obtenido utilizando el método propuesto y mediante el método basado en la *Gaussiana Simple* han sido los siguientes:

	Sin mejora	Con mejora
TP	26096	33438
FN	25149	17807
OTN	362757	362221
FP	718	1254
P0	0.935	0.953
P1	0.973	0.964
R0	0.998	0.998
R1	0.509	0.653
FS0	0.966	0.974
FS1	0.669	0.778
SUMA	1.634	1.753

Tabla 18: Resultados para la secuencia Baile Multimodal

En el caso de una secuencia multimodal como la secuencia BAILE MULTIMODAL se observa como también en este caso mantenemos unos valores aproximados a los obtenidos en la segmentación mediante la *Gaussiana Simple* para la detección del fondo, pero obtenemos unos resultados considerablemente más precisos para la detección del frente.

En la gráfica siguiente observamos la comparativa entre el método sin mejora y el método propuesto:

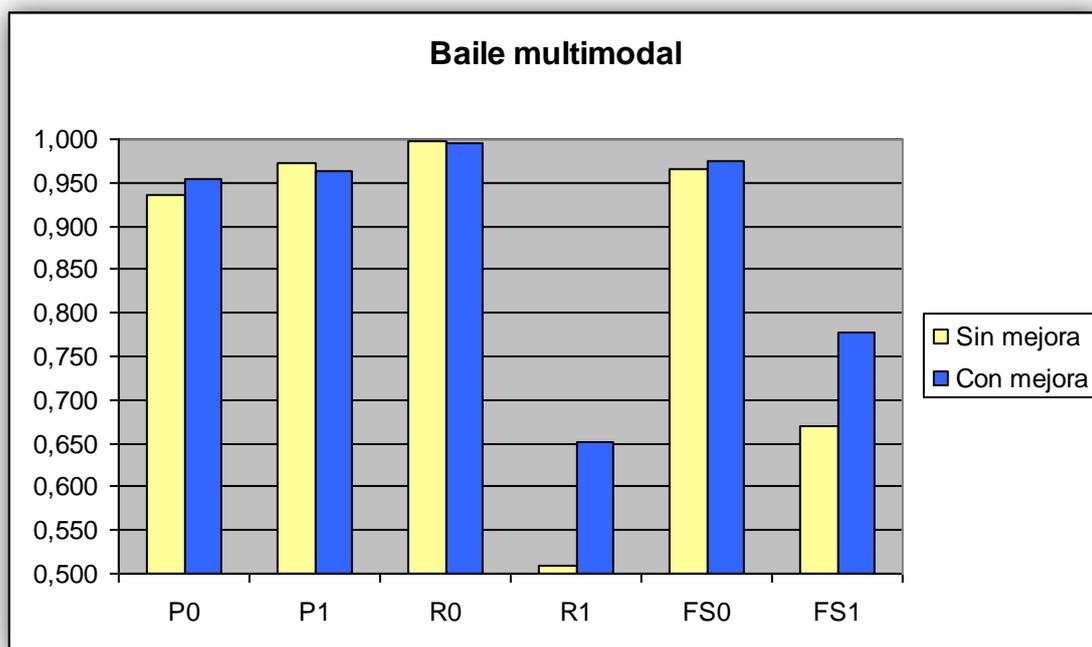


Figura 48: Diagrama de barras de los parámetros estudiados en la secuencia Baile Multimodal

Para el caso del fondo, obtenemos un valor de la precisión (P0) mayor, lo que significa que se detectan la mayoría de los puntos de fondo correctamente con respecto a los que se detectan de manera errónea. Por otra parte, en el valor de *recall* experimentamos una igualdad en los resultados con respecto al método sin mejora.

Para el caso del frente, se observa que si bien el valor de P1 disminuye, es decir, el porcentaje de valores de frente detectados correctamente con respecto a los detectados erróneamente es menor, el valor del *recall* aumenta considerablemente. Este valor muestra que la proporción de valores de frente detectados con respecto a los que realmente son puntos pertenecientes al frente en la máscara *ground truth* aumenta de manera muy significativa, concretamente se mejora un 14,4%.

Los valores de *FScore* que resumen la calidad del sistema recogen una mejoría muy significativa tanto para los valores del frente como del fondo. Obtenemos una mejora de un 10,9% para la precisión de los valores del frente y un 0,8% para el

caso de los valores del fondo, llegando al valor de 97,4% de precisión, prácticamente el 100%.

A continuación se observa una comparativa entre una primera máscara obtenida por el método basado en la *Gaussiana Simple* y una segunda, obtenida mediante el método propuesto.



Figura 49: A la izquierda se muestra la máscara obtenida mediante el método basado en la *Gaussiana Simple* y a la derecha la obtenida con el método propuesto.

En las máscaras obtenidas por ambos métodos obtenemos una calidad muy alta en la detección de las siluetas, sobre todo teniendo en cuenta que estamos ante un escenario multimodal, en el que hay presentes elementos que arrojan un alto nivel de ruido a las imágenes de la secuencias. Pese a ello, se observa como las siluetas se han perfilado y se obtienen zonas de la imagen que previamente no habían sido detectadas.

También se observa como el número de regiones pertenecientes a los objetos ajenos al frente ha sido disminuída considerablemente, y si bien aparecen zonas erróneas ha sido debido a que el estudio del movimiento indica que se mueven de manera coherente.

SECUENCIA 3: TENIS

Los resultados que se han obtenido utilizando el método propuesto y sin utilizarlo han sido los siguientes:

	Sin mejora	Con mejora
TP	8022	11804
FN	23701	19919
TN	381995	378867
FP	1002	4130
P0	0.942	0.950
P1	0.889	0.741
R0	0.997	0.989
R1	0.253	0.372
FS0	0.969	0.969
FS1	0.394	0.495
SUMA	1.362	1.465

Tabla 19: Resultados para la secuencia Tennis

La secuencia Tennis es una secuencia multimodal media, que presenta un nivel de ruido de fondo elevado. Este nivel de ruido viene dado por la cantidad de objetos ajenos al frente que aparecen en la imagen, como pueden ser las ramas de los árboles cercanos. En este caso se observa que se mantiene un nivel de calidad alto, debido a que los resultados obtenidos mediante la segmentación por medio de la *Gaussiana Simple* tienen un nivel elevado de precisión

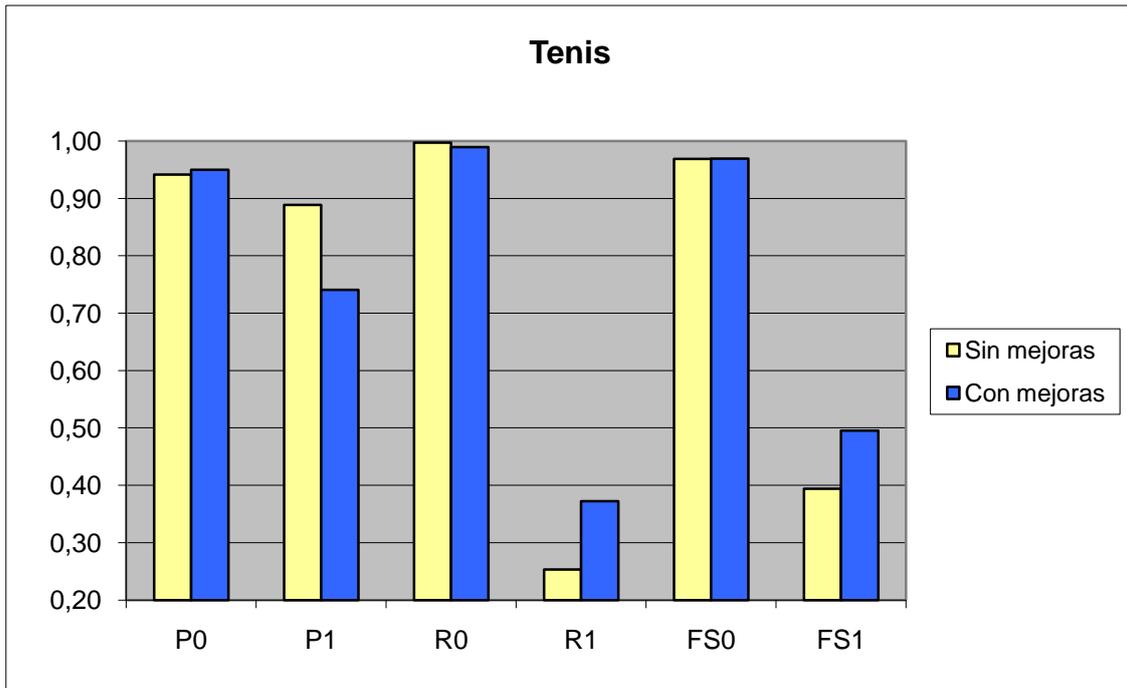


Tabla 20: Diagrama de barras de los parámetros estudiados en la secuencia Tenis.

Para el caso de la detección de los puntos del fondo se observa un ligero aumento en la precisión en cero (P0), que se traduce en un aumento de la proporción de puntos correctamente detectados con respecto a los detectados de forma errónea. También se observa una ligera disminución en cuanto al nivel de *Recall* en ceros. Aún así, el valor alcanzado mediante el método propuesto prácticamente alcanza el 100% de los puntos del fondo detectados con respecto a los encontrados en la máscara *ground truth*, alcanzando un valor de 98,9%.

Para los puntos pertenecientes al frente, se produce un claro descenso en el valor de P1, lo que indica que se han detectado una cantidad mucho menor de puntos correctos comparados con los puntos incorrectos que se han detectado. Este descenso en el nivel de la precisión se ve compensado con la calidad de los puntos detectados. En este caso se produce un claro aumento del nivel de *Recall* de unos. En este caso se produce un aumento de un 11,9% en la calidad de los puntos detectados como frente con respecto a los valores reales de frente, detectados en la máscara *ground truth*. En este caso, debido a la calidad de la segmentación que contábamos como partida, los resultados sólo se han podido mejorar hasta cierto punto.

Teniendo en cuenta las medidas de la calidad total del sistema, tanto para la detección del fondo como para la detección del frente se produce un aumento de la calidad obtenida. Para los valores de frente se obtiene un aumento de un 10,1%, mientras que para los valores del fondo se obtiene exactamente el mismo nivel de precisión.

En las siguientes máscaras, se puede comprobar cómo se han mejorado con el método propuesto.



Tabla 21: A la izquierda se muestra la máscara obtenida mediante el método basado en la Gaussiana Simple y a la derecha la obtenida con el método propuesto.

En las máscaras obtenidas para los dos métodos se observa cómo se ha producido un ligero aumento en la superficie de la silueta, así como se han reducido algunas de las zonas consideradas como frente en la máscara inicial. Por el contrario, al igual que sucedería en la secuencia Baile Unimodal, se han detectado zonas de la imagen que al principio no habían sido detectadas como frente y ahora sí se han detectado, erróneamente, como es el caso de algunas de las zonas del árbol que aparece en el fondo de la secuencia. Esto se debe a que el estudio del movimiento ha asignado a las ramas un movimiento coherente, con lo que para nuestro algoritmo sí se puede considerar como frente.

SECUENCIA 4: INGRAVIDEZ

Los resultados que se han obtenido utilizando el método propuesto y sin utilizarlo han sido los siguientes:

	Sin mejora	Con mejora
TP	13352	22124
FN	52435	43663
TN	346228	346659
FP	2705	2274
P0	0.868	0.888
P1	0.832	0.907
R0	0.992	0.993
R1	0.203	0.336
FS0	0.926	0.938
FS1	0.326	0.491
SUMA	1.253	1.428

Tabla 22: Resultados para la secuencia Ingravidéz

La secuencia Ingravidéz es una secuencia multimodal alta en la que existe un nivel de ruido muy alto en la zona del fondo de las imágenes. Es una secuencia, por tanto, difícil de segmentar con los métodos tradicionales. Con el método desarrollado se consigue una gran mejora a la hora de detectar correctamente los puntos pertenecientes al frente.

En esta secuencia, obtenemos una mejora tanto en la detección de frente como en la detección del fondo. Esto se debe a que el método desarrollado responde de la mejor manera ante situaciones en las cuales el ruido de fondo presenta dificultades a la hora de discernir si un punto pertenece al fondo o al frente. En este caso, la precisión de la detección de los puntos tanto del frente como del fondo aumenta.

En la gráfica siguiente observamos la comparativa entre el método sin mejora y el método propuesto:

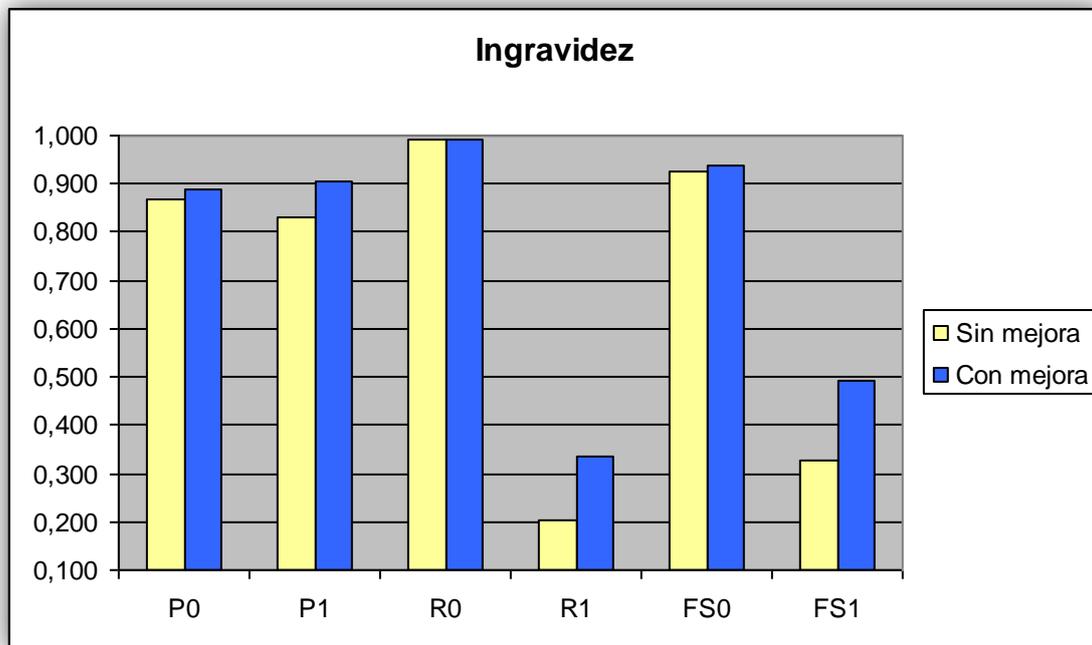


Figura 50: Diagrama de barras de los parámetros estudiados en la secuencia Ingravidez.

En la gráfica observamos que se producen mejoras en todos los parámetros de medición de la calidad.

Para el caso del fondo, se observa como la precisión P0 de los puntos correctamente detectados como fondo con respecto a los puntos detectados erróneamente como fondo aumenta en un 2%. El valor que se obtiene para el *recall* en este caso roza el 100%, lo que indica que prácticamente todos los puntos del fondo han sido detectados.

Con respecto al frente, en este tipo de secuencias es muy difícil obtener un porcentaje tan alto como para el fondo, debido a la dificultad que entraña el tener un constante ruido provocado por objetos externos al frente. Pese a las dificultades intrínsecas de la secuencia, con el método propuesto se consigue realizar una mejora interesante en la detección del frente. En este caso, obtenemos un aumento del 7,5% en la precisión de los valores a uno de la imagen, que se traduce en un

aumento de los puntos detectados correctamente como frente respecto a los detectados erróneamente como frente. El valor del *recall* muestra que se produce una mejora del 13,3% en la detección de puntos del frente, lo que significa que cerca de un 15% de puntos del frente que no habían sido detectados han sido identificados gracias al método propuesto.

Los valores de *FScore* aportan una robusta conclusión con respecto a la mejora que el método ha provocado en la detección de la máscara. Se obtienen valores de precisión un 16,5% mejores para la detección del frente y del 1,2% para el caso de la detección del fondo.

En las siguientes máscaras, se puede comprobar cómo se han mejorado con el método propuesto.

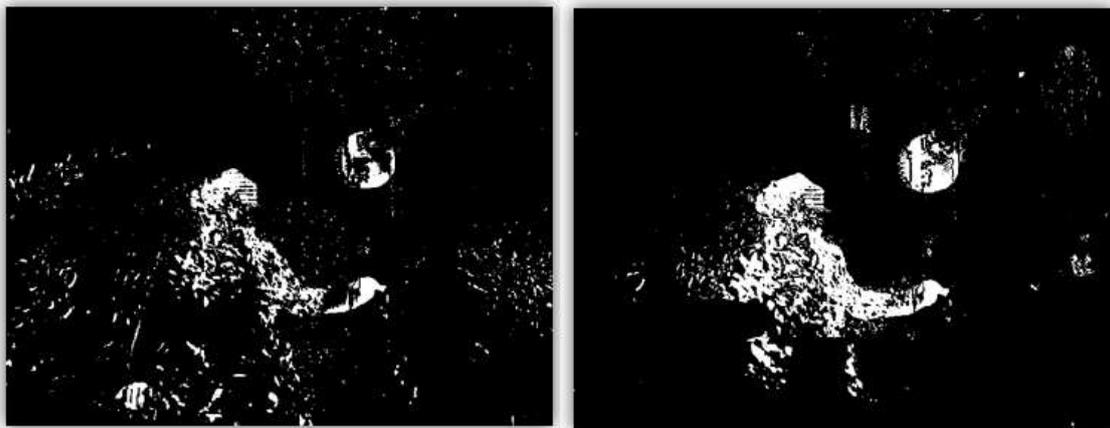


Figura 51: A la izquierda se muestra la máscara obtenida mediante el método basado en la *Gaussiana Simple* y a la derecha la obtenida con el método propuesto.

Se puede observar cómo se han detectado muchas más zonas de la superficie de la persona en cuestión, mediante el agrupamiento correcto de las zonas consideradas como fondo en primera instancia, pero corregidas como frente al realizar el estudio del movimiento. Estas zonas se concentran mayoritariamente en el torso de la silueta, debido a que el principal movimiento que se realiza es un movimiento de flexión en el tronco superior. También observamos cómo la pelota que aparece en las imágenes se completa con un número mayor de regiones.

En las zonas del fondo también se observa la mejoría que se produce, eliminando considerablemente el ruido provocado en este caso por un entorno en el que hay árboles moviéndose por efecto del viento. Esto se explica gracias al proceso de selección de regiones que se ha realizado, en el cual solo aquellas que presentan un movimiento coherente han sido incluidas en la máscara final.

SECUENCIA 5: SILLA

Los resultados que se han obtenido utilizando el método propuesto y sin utilizarlo han sido los siguientes:

	Sin mejora	Con mejora
TP	25374	24597
FN	21988	22765
TN	362843	366054
FP	4515	1304
P0	0.943	0.941
P1	0.849	0.950
R0	0.988	0.996
R1	0.536	0.519
FS0	0.965	0.968
FS1	0.657	0.671
SUMA	1.622	1.640

Tabla 23: Resultados para la secuencia Silla

La secuencia denominada SILLA es una secuencia multimodal alta. Esto indica que el nivel de ruido existente en las imágenes va ser alto y va a dificultar considerablemente la extracción del fondo. En este caso, también obtenemos una mejora considerable de los resultados con respecto al método basado en la *Gaussiana Simple* tanto en la precisión de la detección del frente como del fondo.

En la gráfica siguiente observamos la comparativa entre el método sin mejora y el método propuesto:

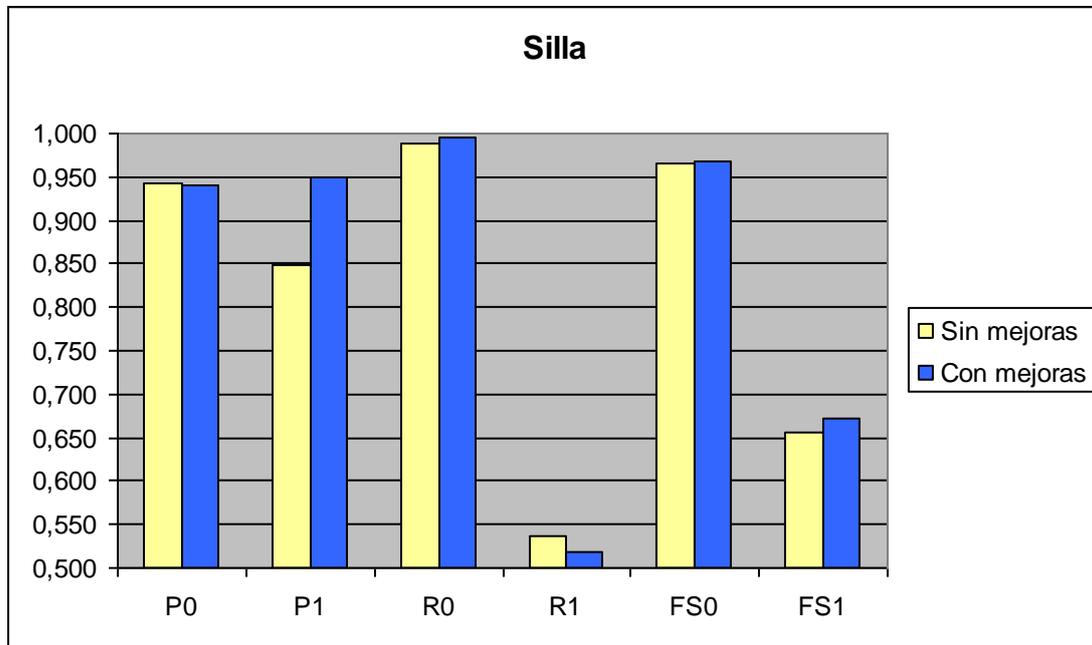


Figura 52: Diagrama de barras de los parámetros estudiados en la secuencia Silla.

Pese a que se produce un ligero descenso de apenas un 0,2% del valor de la precisión P0 (la proporción de puntos de fondo detectados correctamente con respecto a los detectados incorrectamente), el valor del *recall* muestra que la calidad de la detección del fondo ha sido mayor mediante la utilización del método propuesto, aumentado en un 0,8%, y situándose en el 99,6%. Este valor significa que prácticamente la totalidad de los puntos detectados como fondo han sido detectados correctamente.

Como se ha venido comentando a lo largo del tema, la verdadera mejora del método propuesto se observa fundamentalmente en la detección de los puntos considerados como frente. En esta secuencia también se cumple dicha conclusión. Se obtiene primeramente una mejora de un 10,1% en el nivel de precisión de los píxeles detectados como frente correctamente con respecto a los detectados erróneamente como frente. Por otra parte, y contradiciendo el resultado visual, se produce un cierto descenso en el valor obtenido de *recall*. En este caso, se obtiene

un 1,7% de píxeles de frente detectados con respecto a los píxeles que realmente pertenecen al frente en la máscara *ground truth*.

Los valores que resumen la calidad de la segmentación denominados *FScore*, también muestran un aumento de la calidad del método propuesto. En este caso se obtiene un aumento de un 1,4% en el caso de los puntos del frente, y un 0,3% para el caso de los puntos pertenecientes la fondo.

A continuación se observan dos máscaras obtenidas por cada uno de los dos métodos, en las cuales se observan las notables diferencias entre el sistema sin mejora y el sistema con mejora.

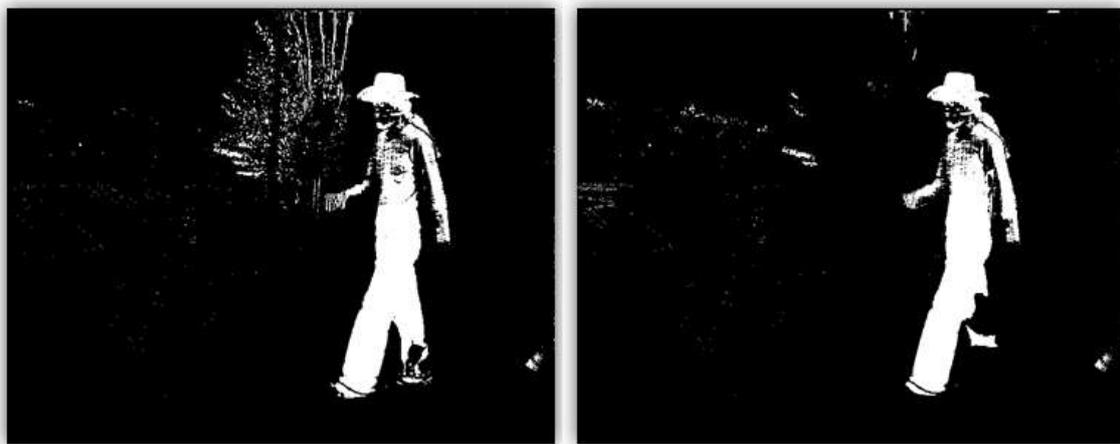


Figura 53: A la izquierda se muestra la máscara obtenida mediante el método basado en la *Gaussiana Simple* y a la derecha la obtenida con el método propuesto.

En esta caso se observa muy claramente cómo se ha completado la figura de la persona presente en la secuencia, salvo en algunas zonas de las piernas, en las cuales no aparecen zonas que previamente sí se obtenían. En cambio, se produce una mejora considerable de las zonas que previamente constituían el frente y que gracias al estudio que se ha realizado han sido eliminadas del mismo. En este caso, se observa cómo en la máscara obtenida mediante el método basado en la *Gaussiana Simple* aparecen ramas de un árbol, cuyo movimiento a priori parecía presagiar que fuera parte del frente. Sin embargo, después de un estudio

exhaustivo, se considera que la mayoría de las zonas establecidas previamente no eran en realidad zonas de interés.

5.6.1.2. Resultados globales

Para comparar con mayor claridad los resultados de los algoritmos con y sin la mejora realizada expuestos en los apartados anteriores se han realizado diagramas en los que se compara cada uno de los parámetros para todas las secuencias estudiadas. Esta comparación mostrará una visión general de los resultados obtenidos, que servirá para obtener las conclusiones sobre la aplicación del método propuesto para cada una de las secuencias. Los resultados son los que se exponen a continuación:

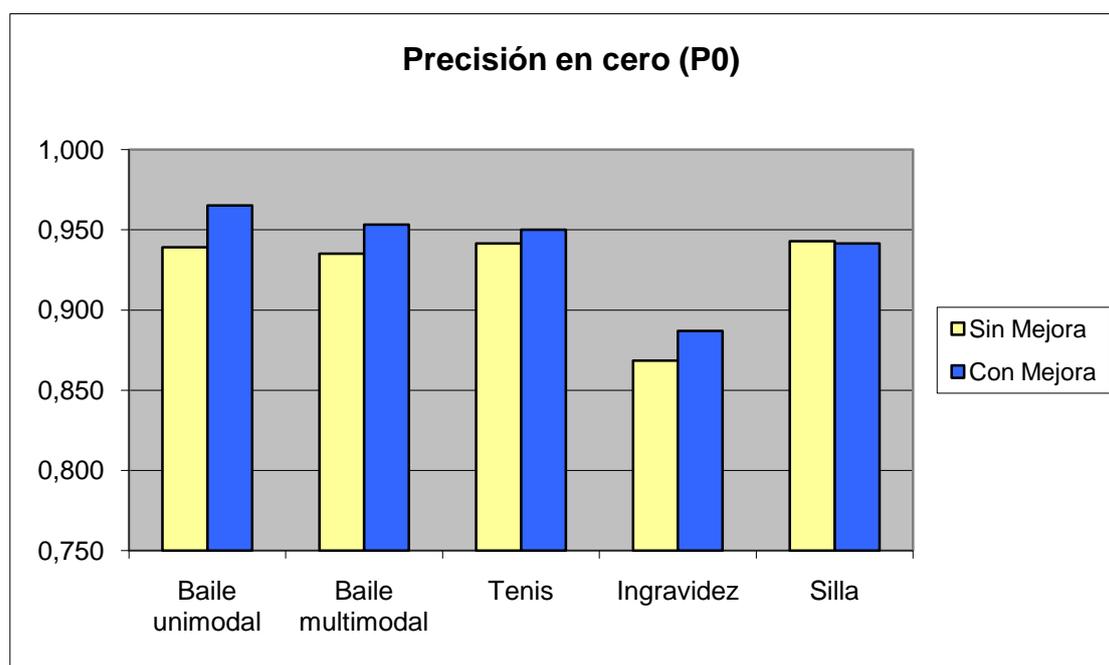


Figura 54: Diagrama de barras de la precisión en cero por cada secuencia procesada para los dos métodos.

En cuanto al nivel de la precisión en cero, se obtiene un aumento general de la calidad, con la única excepción de una muy ligera disminución en el caso de la secuencia Silla (en la que el valor desciende un 0,1%). Para todos los demás casos,

se ha producido un aumento de la precisión P0. Este valor indica la proporción de los puntos detectados correctamente como fondo con respecto al número total de puntos de fondo detectados (la suma de los detectados correctamente e incorrectamente). El método, por lo tanto, aumenta significativamente, hasta un 2,6%, la calidad de la detección del fondo de la secuencia.

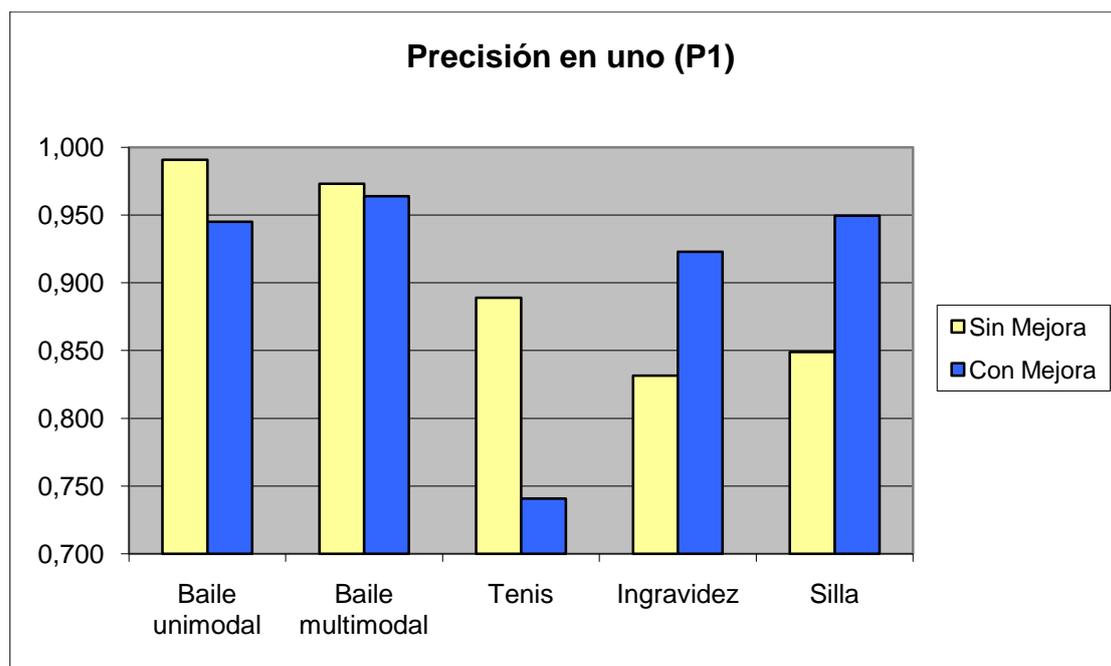


Figura 55: Diagrama de barras de la precisión en uno por cada secuencia procesada para los dos métodos

La precisión en uno detectada presenta resultados diferentes dependiendo de las características de la secuencia estudiada. Se observa como para las secuencias Baile Unimodal y Baile Multimodal se produce un ligero descenso de la precisión, debido a que se han detectado un mayor número de puntos como frente tanto correctamente como erróneamente. Esta disminución es una pérdida de calidad relativa, debido a que es una comparativa entre los puntos de frente detectados correctamente respecto al total. Esto puede significar que la calidad del sistema puede verse mejorada pese a haber obtenido un valor de P1 menor. En el caso de la secuencia Tennis, sí se observa un claro descenso entre los valores obtenidos mediante el método de la *Gaussiana Simple* y los obtenidos mediante el método propuesto. Para las secuencia Ingravidez y Silla, los valores aumentan

significativamente, debido principalmente a que se detectan menos puntos erróneos de frente, que en la comparativa supone una mejora de la precisión.

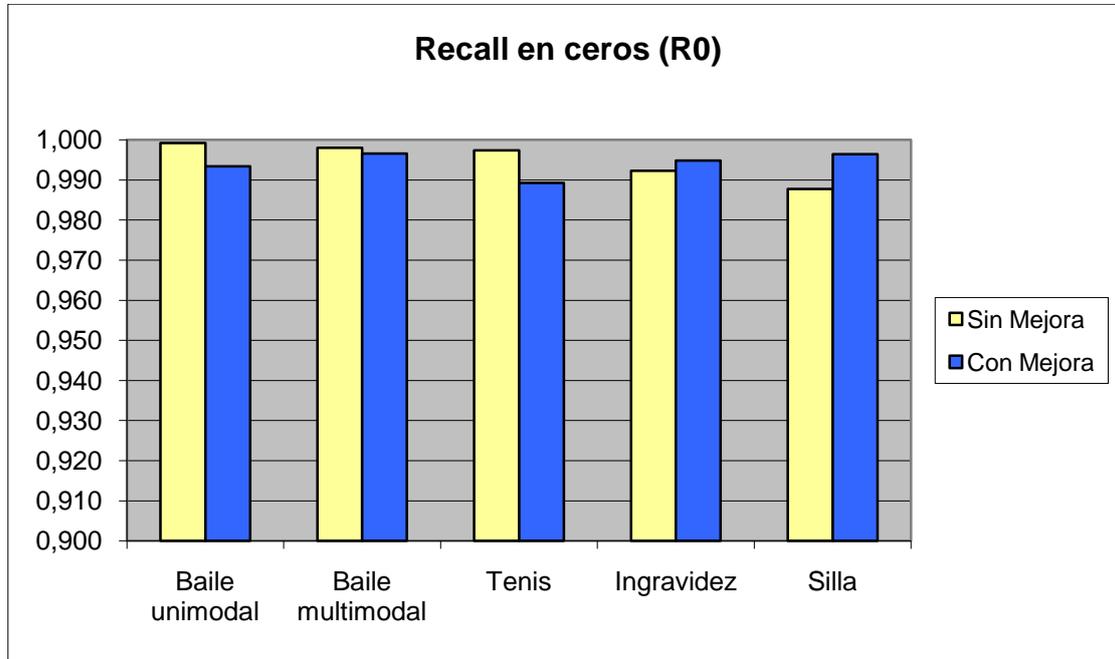


Figura 56: Diagrama de barras del *recall* en ceros por cada secuencia procesada para los dos métodos.

El valor del *Recall* en cero nos muestra el grado de precisión en la detección del fondo de la imagen. Este valor muestra el grado de fidelidad con el fondo calculado en la máscara *ground truth*. Los valores obtenidos muestran unos resultados que rozan la perfección en la detección correcta del fondo. En el método propuesto se ha centrado el interés en la detección correcta de los píxeles del frente, y es donde se encuentran los resultados más precisos. Para el caso del *recall* en cero, prácticamente se mantiene una calidad cercana al 100% que es similar a la obtenida mediante el método basado en la *Gaussiana Simple* y que sin lugar a dudas es muy difícil superar. Para todos los métodos obtenemos una calidad cercana al 99% de precisión.

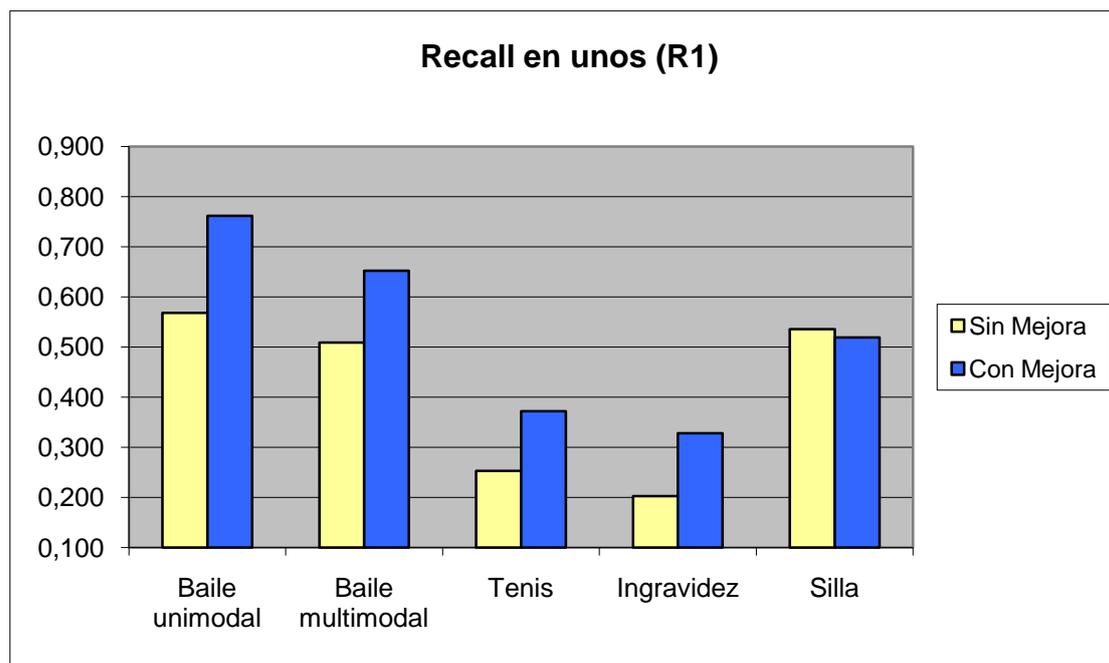


Figura 57: Diagrama de barras del recall en unos por cada secuencia procesada para los dos métodos.

En el caso del *Recall* en unos, obtenemos unos resultados que mejoran considerablemente los obtenidos con el método sin mejora. Esto se aprecia en todas las secuencias estudiadas salvo en la secuencia Silla, en la cual se produce un descenso del 1,4% en la calidad. Dadas las características del algoritmo empleado en el método propuesto, la mejora en la obtención del frente es donde se obtienen los resultados más precisos en comparación con los métodos de segmentación existentes en el estado del arte. Esto se debe a que partiendo de un método muy preciso, se realiza una mejora que surte efecto en la inmensa mayoría de las secuencias analizables.

En esta gráfica se observa cómo se aumenta la precisión de los puntos de la imagen detectados como frente con respecto a los que realmente son frente en la máscara de *ground truth*. La mejora llega a obtener un 20% de mejoría en la detección de puntos del frente con respecto al método sin mejora. Es decir, se obtiene un 20% más de frente que en el resto de métodos del estado del arte.

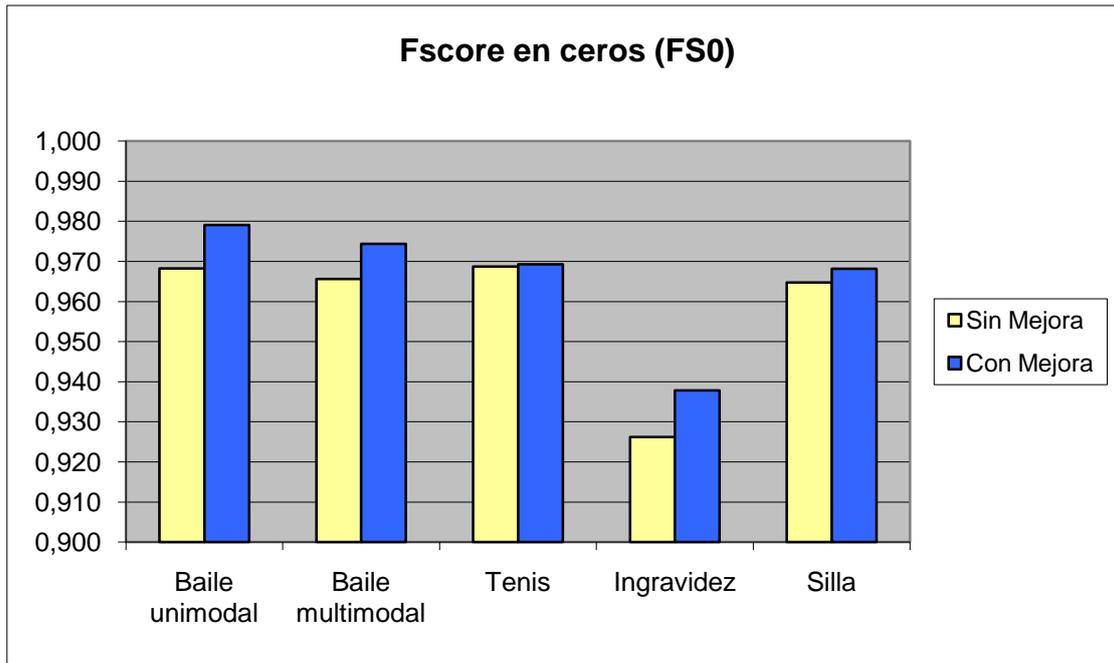


Figura 58: Diagrama de barras del *Fscore* en ceros por cada secuencia procesada para los dos métodos.

El valor de *Fscore* muestra una medida de calidad en la que se unen la precisión obtenida con el *recall*. Tenemos por tanto una medida muy fiable de la calidad en la detección de, en este caso, el fondo. Como muestra la figura 63, se obtiene una mejora en la calidad de la detección del fondo en todas las secuencias estudiadas. Esto significa que con el método propuesto se consigue mejorar la segmentación del fondo. Los valores muestran en todos los casos salvo en la secuencia Ingravidéz, vales cercanos al 97% de fiabilidad del sistema, lo que hace al método propuesto, una alternativa muy interesante en la segmentación de objetos en movimiento.

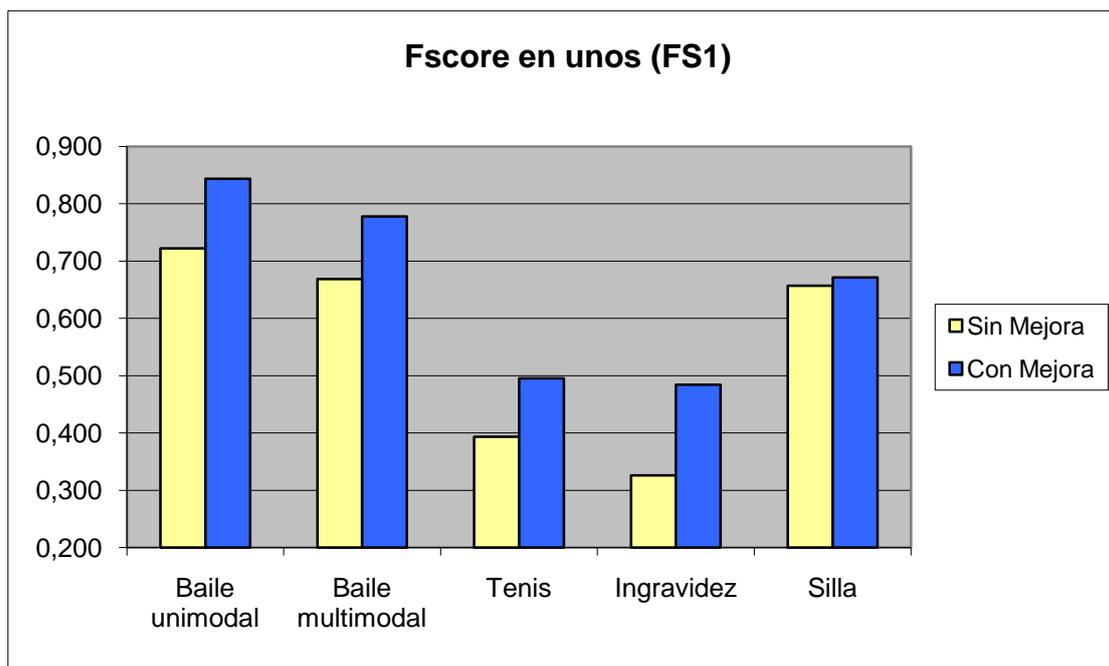


Figura 59: Diagrama de barras del *Fscore* en unos por cada secuencia procesada para los dos métodos.

En la gráfica anterior observamos cómo con el método propuesto obtenemos un gran mejora de la calidad para todas las secuencias estudiadas. En la mayoría de ellas se obtienen mejoras de más del 12% en la calidad de la detección, llegando incluso a obtenerse un 17% de mejora en la calidad del sistema. Nuevamente se pone de manifiesto que con el algoritmo empleado, se consigue de una manera sencilla mejorar uno de los mejores métodos de segmentación existentes en el estado del arte. Y es en la detección de los puntos del frente donde el sistema propuesto obtiene los mejores resultados.

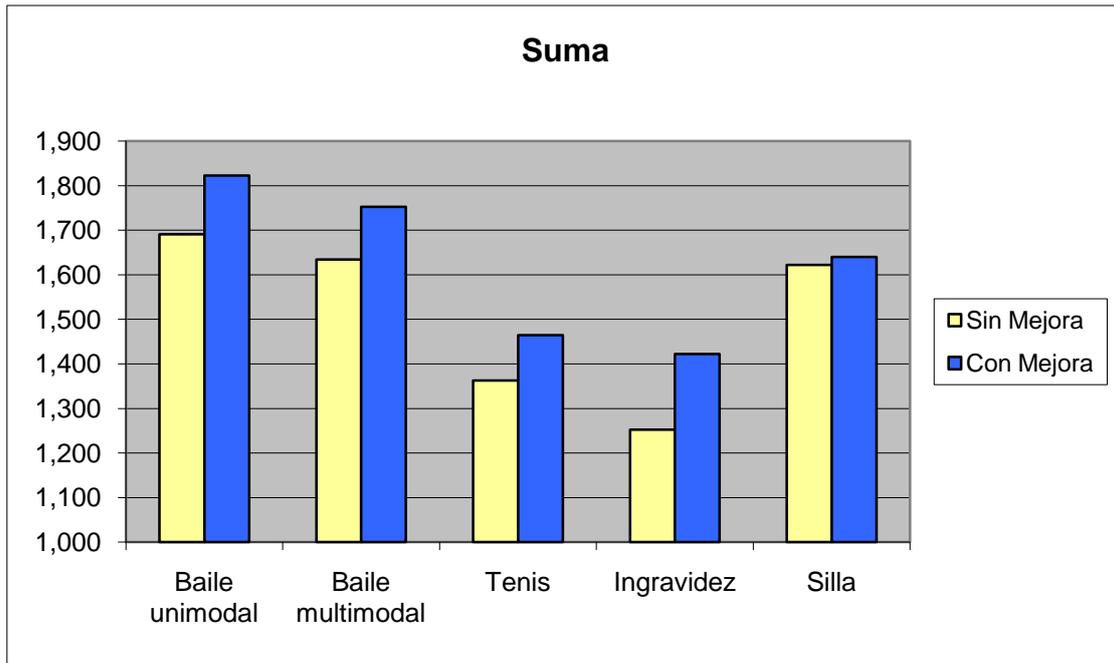


Figura 60: Diagrama de barras de la suma en unos por cada secuencia procesada para los dos métodos.

Para mostrar un valor final que resuma las capacidades de los sistemas estudiados, utilizamos el parámetro *suma*, que consiste en englobar los parámetros anteriores en un sólo valor. Como conclusión se muestra como **todas las secuencias estudiadas han sido mejoradas por medio del método propuesto.**

5.6.1.3. Resultados: Conclusiones

Llegado a este punto, se hará un breve resumen de las conclusiones obtenidas tras la visualización de los resultados obtenidos. Se debe resaltar cómo mediante un método que ha sido diseñado adaptándose a las circunstancias y a los resultados que se iban obteniendo, se ha alcanzado un nivel de mejoría importante con respecto al método más preciso existente en el estado del arte. En los resultados que se han ido explicando debemos hacer hincapié en la mejora que se obtiene de la detección de los objetos del frente, y como se ha visto en la secuencias Silla, se pueden eliminar las zonas de error del fondo mediante un estudio exhaustivo de su movimiento. Este método, por tanto, resulta útil para realizar la segmentación.

6

CONCLUSIONES Y TRABAJO FUTURO

En este PFC se ha desarrollado un método que realice una mejora de los métodos de segmentación existentes en el estado del arte. Para ello, se ha realizado un estudio exhaustivo de la influencia que la información de movimiento puede tener a la hora de realizar una segmentación más precisa.

La técnica de segmentación escogida para realizar la máscara inicial del estudio ha sido una técnica paramétrica de segmentación. Concretamente la técnica basada en la Gaussiana Simple. Esta técnica ha sido escogida debido a la fiabilidad que presentaban sus resultados, que servían de punto de partida para el método desarrollado en este proyecto.

Las técnicas de estimación de movimiento que han sido estudiadas se han dividido en métodos basados en el gradiente (Flujo óptico) y métodos basados en desplazamiento de bloques (todos los métodos de Block Matching existentes en el estado del arte). Con ellos, se ha realizado una comparación de sus características y

capacidades y se ha seleccionado el método que mejores resultados ofrecía para el desarrollo planteado. Este método ha sido el estudio del flujo óptico.

Posteriormente, se han introducido los conceptos de los métodos de agrupamiento, y dentro de los algoritmos existentes, se ha optado por la utilización del algoritmo de Mean Shift. Con este método se han evaluado las condiciones de los vectores de movimiento extraídos en la estimación.

Finalmente, para evaluar el método que se ha realizado, se han realizado diversas medidas objetivas y se han realizado gráficas, diagramas de barras y tablas para llegar a las conclusiones que se exponen a continuación.

La segmentación de objetos en movimiento sigue siendo un reto hoy en día en el procesamiento de secuencias complejas (secuencias multimodales altas), debido a que los algoritmos de segmentación basados en el modelado de fondo diseñados para este fin no son capaces de extraer un objeto en movimiento correcto si los píxeles del fondo presentan una variación más significativa que los del mismo objeto.

Por ello, se hace cada vez más importante la integración de otros métodos en la segmentación. Estos métodos pueden ser los recogidos en este proyecto, como puede ser el estudio del movimiento que sufren los píxeles de la imagen, mediante el cual se puede obtener una información que nos resulte útil a la hora de discernir si un determinado objeto pertenece o no al frente de la imagen.

Para que el método que se ha implementado resulte preciso, es muy importante realizar un estudio de todos los parámetros que se utilizan a lo largo del proceso, con el fin de obtener resultados fiables para todo tipo de secuencias de video.

6.1. Trabajo futuro

La integración de la información de movimiento en la segmentación es un campo que ofrece multitud de posibilidades por investigar: los modelos matemáticos que realizan la estimación de movimiento; nuevos métodos de estimación basados, por ejemplo en el estudio de puntos característicos; nuevos métodos de agrupamiento, etc...

Una de las posibles líneas de trabajo futuro podría ser el estudio del movimiento de los píxeles de la imagen en el tiempo, lo que se conoce como *tracking*. Este proceso resolvería los problemas introducidos por el movimiento de los objetos del fondo en la segmentación. Sería interesante un estudio de cada uno de los píxeles de la imagen a lo largo del tiempo, almacenando el movimiento que experimentan en una secuencia completa. De esta manera, se puede obtener un estudio preciso de las características del movimiento de todas las regiones de la imagen, y mediante métodos de selección discernir cuál de ellas pertenece realmente a objetos de interés.

Por otro lado, la integración de la información de movimiento en la segmentación necesita métodos robustos en cada una de las etapas para la obtención de resultados precisos. En particular los métodos de estimación de movimiento. Estos métodos tienen por regla general una alta o muy alta carga computacional que impide que a día de hoy sean utilizados para aplicaciones en tiempo real. Esto se podría evitar mediante la utilización de métodos de estimación de movimiento basados en puntos característicos, que únicamente muestren el movimiento en determinadas zonas o puntos de la imagen, reduciendo considerablemente la carga computacional a cambio de obtener menos información de movimiento.

El tiempo de procesado de los algoritmos también podría reducirse considerablemente utilizando otros lenguajes de programación para la ejecución del método, como podrían C o C++, el lugar de la utilización del lenguaje Matlab, considerablemente más costoso.

BIBLIOGRAFÍA Y REFERENCIAS:

1. **Wang, Yao, Ostermann, Jörn y Zhang, Ya-Quin.** *Video Processing and Communications*. s.l. : Prentice Hall, 2002.
2. *True Motion Estimation - Theory, Application and Implementation.* **Chen, Ywn-Kuang.** Princeton University : s.n., 1998.
3. *Motion Estimation for Region-Based Video Coding.* **Dang, Viet-Nam, Mansouri, Abdol-Reza y Konrad, Janusz.** Institut National de la Recherche Scientifique, Canada : s.n.
4. *Video Object Segmentation by 2-D Mesh-Based Motion Analysis.* **Wang, Yujian y Wu, Zhenyang.** International Conference on Semantic Computing.
5. **Torr, P.H.S y Murray, D.W.** Statistical detection of independent movement from a moving camera. *Image and Vision Computing*. 1993.
6. **Carpio, Jorge Badenas.** Estimación y segmentación de movimiento basado en regiones y su aplicación a tareas de monitorización de tráfico. *Tesis doctoral*. s.l. : Universitat Jaume I.
7. *Measurement of Image Velocity.* **Fleet, D.J.** s.l. : Kluwer Academic Publishers, 1992.
8. *Scene Segmentation from visual motion using global optimization.* **Murray, D.W. y Buxton, B.F.** s.l. : IEEE Transactions on Pattern Analysis and Machine Intelligence, 1989, Vol. 9.
9. *Principles and Practice of Background Maintenance.* **T., Kentaro, K., John y Wallflower, Brian.** s.l. : ICCV. Seventh Int. Conf. on Computer Vision, 1999.
10. *Background and foreground modeling using nonparametric kernel density estimation for visual surveillance.* **Elgammal, A., y otros.** s.l. : IEEE, 2002.
11. *Foreground Segmentation Using Adaptive Mixture Models in color and Depth.* **Harville, M., Gordon, G. y Woodfill, J.** s.l. : EVENT'01: IEEE Workshop on Detection and Recognition of Events in Video, 2001.
12. *A Ground Truth for Motion-Based Video-Object Segmentation.* **Tiburzi, Fabrizio, y otros.** s.l. : Grupo de Tratamiento de Imágenes, Escuela Politécnica Superior, Universidad Autónoma de Madrid.
13. **Martín, Sonsoles Herrero.** Análisis comparativo de técnicas de segmentación de secuencias de video basadas en el modelado de fondo. 2009.
14. **Cheung, S.C. y Kamath, C.** Robust techniques for background subtraction in urban traffic video. *Proc. Elect. Imaging: Visual Comm. Image Processing*. s.l. : S. Panchanathan and B. Vasudev, editors, 2004.
15. *Image change detection algorithms: A systematic survey.* **Radke, R.J., y otros.** March 2005. *Image Processing*. Vol. 14, pág. 294:307.
16. *Multiple view geometry in computer vision.* **Hartley, R. y Zisserman, A.** s.l. : Cambridge University Press, 2000.
17. *Multi-level background initialization using Hidden Markov Models.* **Cristani, M., Bicego, M. y Murino, V.** 2003. First ACM SIGMM Int. workshop on Video surveillance. págs. 11-20.
18. *Frame difference normalization: an approach to reduce error rates of cut detection algorithms for MPEG videos.* **Ewerth, Ralph y Freisleben, Bernd.** 2003. *ICIP* (2). págs. 1009-1012.
19. *Learning Patterns of activity using real time Tracking.* **C., Stauffer y W., Grimson.** Aug 2000. *IEEE Trans. on Pattern Analysis and Machine Intelligence*. págs. 22:747-57.

20. *Stationary background generation: An alternative to the difference of two images.* **Long, W. y Yang, Y.H.** 1990. *Pattern Recognition*. Vol. 23 (12), págs. 1351-1359.
21. *From tracking to advanced surveillance.* **Fuentes, L. y Velastin, S.** Barcelona (Spain) : s.n., Sept. 2003. *IEEE Conference on Image Processing*.
22. *Illumination independent change detection for real world omage sequences.* **Skifstad, K. y Jain, R.** June 1989. *Comp. Vision, Graphics Image Process.* Vol. 46, págs. 387-399.
23. **Jain, R., Kasturi, R. y Schunk, G.B.** *Machine Vision*. s.l. : McGraw Hill Int. Editions, 1995.
24. *Robust and illumination invariant change detection based on linear dependence for surveillance applications.* **Durucan, E. y Ebrahimi, T.** Tampere, Finland : s.n., 2000. *Proc. of X EUSIPCO*. págs. 1041-1044.
25. *Bayesian spatio-temporal motion detection under varying illumination.* **Toth, D., Aach, T. y Metzler, V.** Tampere, Finland : s.n., 2000. *Proc. of X EUSIPCO*. págs. 3-7.
26. *Automatic moving object and background separation.* **Neri, A., y otros.** 1998. Vol. 66 (2), págs. 219-232.
27. *Detection and description of moving objects by stochastic modelling and analysis of complex scenes.* **Hotter, M., Mester, R. y Muller, F.** 1996. *Signal Proces: Image Comm.* Vol. 8, págs. 281-293.
28. **Sonka, M., Hlavac, V. y Boyle, R.** *Image Processing. Analysis and Machine Vision*. s.l. : Chapman & Hall Computing, 1993.
29. **Holden, M.** *Parallel Algorithms for Digital Image Processing*. s.l. : John Wiley and Sons, 1993.
30. **Thévenaz, P.** *Motion analysis. Pattern Recognition and Image Processing in Physics*. s.l. : Scottish Universities Summer School in Physics, 1990, págs. 129-166.
31. **Lucas, B.D. y Kanade, T.** *An iterative Image Registration Technique with an Application to Stereo Vision. Proceeding of Imaging Understanding Workshop*. 1981, págs. 121-130.
32. **Ñesi, P.** *Variational approach to optical flow estimation managing discontinuities. Image and Vision Computing*. Septiembre 1993., Vol. 11 (7), págs. 419-439.
33. **Enkelmann, W.** *Investigations of multigrid algorithms for estimation of optical flow fields in image sequences. Computer Vision, Graphics and Image Processing*. 1988, págs. 43:150-177.
34. **Heitz, F. y Bouthemy, P.** *Multimodal estimation of discontinuous optical flow using markov random fields. IEEE Transations on Pattern Analysis and Machine Intelligence*. 1993, págs. 15:1217-1232.
35. **Barnard, S.T.** *The image correspondence problem*. s.l. : Universidad de Minnesota, 1979.
36. *Disparity analysis of images.* **Barnard, S.T. y Thompson, W.B.** 1980, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, págs. 333-340.
37. *Matching canny edges to compute the principal components of optic flow.* **Castellow, D.A., y otros.** September 1987. *3rd Alvey Vision Conference*. págs. 193-200.
38. *A computational approach to edge detection.* **Canny, J.F.** 1986, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 8, págs. 679-698.
39. *Correspondence problem in image sequence analysis.* **Lef, C.H. y Joshi, A.** 1993, *Pattern Recognition*, Vol. 26, págs. 47-61.
40. *Computation of component image velocity from local phase information.* **Fleet, D.J. y Jepson, A.D.** 1990, *International Journal of Computer Vision*, págs. 5:77-104.

41. *The estimation of the gradient of a density function, with applications in pattern recognition.* **Fukunaga, K. y Hostetler, L.** 1975. IEEE Transactions on Information Theory. Vol. 21(1), págs. 32-40.
42. *Mean Shift: A robust approach toward feature space analysis.* **Comaniciu, Dorin y Meer, Peter.** 5, May de 2002, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, págs. 603-619.
43. **Nagel, H.H.** Displacement vectores derived from second-order intensity variations in images sequiences. *Computer Graphics and Image Processing.* 1983, págs. 21:85-117.
44. **Mitiche, A., Wang, Y.F. y Aggarwal, J.K.** Experiments in computing optical flow with gradient-based, multiconstraint method. *Pattern Recognition.* June 1987, págs. 20(2):173-79.
45. **Haralick, R.M. y Lee, J.S.** The facet approach to optical flow. *Image Understanding Workshop.* 1993.
46. *Determining optical flow.* **Horn, B.K.P. y Schunck, B.G.** 1981, Artificial intelligence, págs. 17(1-3):185-203.
47. *High Accuracy Optical Flow Estimation based on a Theory of Warping.* **Brox, Thomas, y otros.** Prague : s.n., 2004. 8th European Conference on Computer Vision.
48. *Block Matching Algorithms for Motion Estimation.* **Barjatya, Aroh.** 26 de April de 2004, IEEE Trans. Circuits and Systems For Video Technology.
49. *A New Three Step Search Algorithm for Block Matching Estimation.* **Li, Renxiang, Zeng, Bing y Liou, Ming L.** 1994, IEEE Trans. Circuits And Systems for Video Technology, Vol. 4, págs. 438-442.
50. *A Simple and Efficient Search Algorithm for Block-Matching Motion Estimation.* **Lu, Jianhua y Liou, Ming L.** 1997, IEEE Trans. Circuits And Systems for Video Technology, Vol. 2, págs. 429-433.
51. *A novel Four-Step Search Algorithm for Fast Block Motion Estimation.* **Po, Lai-Man y Ma, Wing-Chung.** 3, June de 1996, IEEE Trans. Circuits and Systems for Video Technology, Vol. 6, págs. 313-317.
52. *A new Diamond Seach Algorithm for Fast Block-Matching Motion Estimation.* **Zhu, Shan y Ma, Kai-Kuang.** 2, Feb. de 2000, IEEE Trans. Image Processing, Vol. 9, págs. 287-290.
53. *Evaluation of video segmentation methods for surveillance applications.* **McKoen, K., y otros.** Sep. 2008. Proc. EUSIPCO. Vol. 2, págs. 1045-1048.
54. *Objetive evaluation of relative segmentation quality.* **Pereira, P. Correia and F.** Sept. 2000. Proc. ICIP. Vol. 2, págs. 308-311.
55. *A GROUND TRUTH for Motion-based Video-object Segmentation.* **Tiburzi, Fabrizio, y otros.** San Diego (USA) : s.n., Oct. 2008. IEEE International Conference on Image Processing, ICIP'2008, Workshop on Multimedia Information Retrieval: New Trends and Challenges. págs. 17-20.

ANEXO 1

En este anexo se presentan unas tablas que contienen los resultados del análisis de los parámetros del método de *Mean Shift* utilizado en la integración de la información de movimiento en la segmentación.

Las tablas recogen los resultados obtenidos para cada una de las secuencias analizadas, y en base a los mismos se han seleccionado los parámetros utilizados en la integración final. Como se ha explicado a lo largo del proyecto, el valor de Suma más alto corresponde a la prueba realizada con mayor calidad.

La variable **BW** que aparece en las tablas corresponde a los tres valores de los anchos de banda utilizados. Es decir: 678 corresponde a Bandwidth1 con valor 6, Bandwidth2 con valor 7 y Bandwidth3 con valor 8.

	BAILE2 CON MEJORAS										
BW	TP	FN	TN	FP	P0	P1	R0	R1	FS0	FS1	SUMA
566	21524	32152	360256	788	0,918	0,965	0,998	0,401	0,956	0,567	1,523
567	18956	35624	358458	1682	0,910	0,918	0,995	0,347	0,951	0,504	1,455
568	19568	35412	350125	9615	0,908	0,671	0,973	0,356	0,940	0,465	1,405
576	21548	32415	360256	501	0,917	0,977	0,999	0,399	0,956	0,567	1,523
577	21245	31145	358025	4305	0,920	0,832	0,988	0,406	0,953	0,545	1,498
578	33256	21245	359854	365	0,944	0,989	0,999	0,610	0,971	0,755	1,726
586	31245	22145	360001	1329	0,942	0,959	0,996	0,585	0,968	0,727	1,695
587	31425	22145	359845	1305	0,942	0,960	0,996	0,587	0,968	0,728	1,697
588	22745	28114	359874	3987	0,928	0,851	0,989	0,447	0,957	0,586	1,544
666	12413	41710	360475	122	0,896	0,990	1,000	0,229	0,945	0,372	1,318
667	14298	39825	360272	325	0,900	0,978	0,999	0,264	0,947	0,416	1,363
668	14326	39797	360265	332	0,901	0,977	0,999	0,265	0,947	0,417	1,364
676	21222	32901	360274	323	0,916	0,985	0,999	0,392	0,956	0,561	1,517
677	33696	20430	359726	868	0,946	0,975	0,998	0,623	0,971	0,760	1,731
678	32895	21228	359697	900	0,944	0,973	0,998	0,608	0,970	0,748	1,718
686	35206	18917	359520	1077	0,950	0,970	0,997	0,650	0,973	0,779	1,752
687	36110	18013	359387	1210	0,952	0,968	0,997	0,667	0,974	0,790	1,764
688	40213	13910	356489	4108	0,962	0,907	0,989	0,743	0,975	0,817	1,792
699	37873	16250	359208	1389	0,957	0,965	0,996	0,700	0,976	0,811	1,787

BAILE MULTIMODAL CON MEJORAS											
BW	TP	FN	TN	FP	P0	P1	R0	R1	FS0	FS1	SUMA
566	24134	27111	363021	454	0,931	0,982	0,999	0,471	0,963	0,637	1,600
567	26832	24413	362985	490	0,937	0,982	0,999	0,524	0,967	0,683	1,650
568	26712	24533	362945	530	0,937	0,981	0,999	0,521	0,967	0,681	1,647
576	28990	22255	362722	753	0,942	0,975	0,998	0,566	0,969	0,716	1,685
577	31415	19830	362730	745	0,948	0,977	0,998	0,613	0,972	0,753	1,726
578	30993	20252	362797	678	0,947	0,979	0,998	0,605	0,972	0,748	1,720
586	32743	18502	362418	1057	0,951	0,969	0,997	0,639	0,974	0,770	1,744
587	32989	18256	362509	966	0,952	0,972	0,997	0,644	0,974	0,774	1,749
588	33922	17523	362135	1140	0,954	0,967	0,997	0,659	0,975	0,784	1,759
588	33225	18020	362240	1235	0,953	0,964	0,997	0,648	0,974	0,775	1,749
666	19283	31962	363262	213	0,919	0,989	0,999	0,376	0,958	0,545	1,503
667	23567	27678	363211	264	0,929	0,989	0,999	0,460	0,963	0,628	1,591
668	23895	27350	363219	256	0,930	0,989	0,999	0,466	0,963	0,634	1,597
676	28807	22438	362376	1099	0,942	0,963	0,997	0,562	0,969	0,710	1,679
677	31222	20023	362399	1076	0,948	0,967	0,997	0,609	0,972	0,747	1,719
678	31036	20209	362221	1254	0,947	0,961	0,997	0,606	0,971	0,743	1,714
686	32003	19239	362182	1296	0,950	0,961	0,996	0,625	0,972	0,757	1,730
687	33225	18020	362240	1235	0,953	0,964	0,997	0,648	0,974	0,775	1,749
688	33604	17641	362065	1410	0,954	0,960	0,996	0,656	0,974	0,779	1,754

TENIS CON MEJORAS											
BW	TP	FN	TN	FP	P0	P1	R0	R1	FS0	FS1	SUMA
566	5596	26127	382566	431	0,936	0,928	0,999	0,176	0,966	0,296	1,263
567	7730	23993	382375	622	0,941	0,926	0,998	0,244	0,969	0,386	1,355
568	11233	20490	379363	3634	0,949	0,756	0,991	0,354	0,969	0,482	1,451
576	9895	21828	381483	1514	0,946	0,867	0,996	0,312	0,970	0,459	1,429
577	9952	21771	381518	1479	0,946	0,871	0,996	0,314	0,970	0,461	1,432
578	9936	21787	381457	1540	0,946	0,866	0,996	0,313	0,970	0,460	1,430
586	11075	20648	379146	3851	0,948	0,742	0,990	0,349	0,969	0,475	1,444
587	7633	24090	382351	646	0,941	0,922	0,998	0,241	0,969	0,382	1,350
588	11136	20587	379401	3596	0,949	0,756	0,991	0,351	0,969	0,479	1,449
666	6545	28781	378342	1052	0,929	0,862	0,997	0,185	0,962	0,305	1,267
667	7869	23123	382112	1616	0,943	0,830	0,996	0,254	0,969	0,389	1,357
668	7140	24712	381288	1580	0,939	0,819	0,996	0,224	0,967	0,352	1,319
676	7864	25637	381123	96	0,937	0,988	1,000	0,235	0,967	0,379	1,347
677	8127	23221	381122	2250	0,943	0,783	0,994	0,259	0,968	0,390	1,357
678	9881	21332	380083	3424	0,947	0,743	0,991	0,317	0,968	0,444	1,412
686	8985	20997	380442	4296	0,948	0,677	0,989	0,300	0,968	0,415	1,383
687	9992	21765	381223	1740	0,946	0,852	0,995	0,315	0,970	0,460	1,430
688	9967	21129	381998	1626	0,948	0,860	0,996	0,321	0,971	0,467	1,438

INGRAVIDEZ CON MEJORAS											
BW	TP	FN	TN	FP	P0	P1	R0	R1	FS0	FS1	SUMA
555	7279	58508	348459	474	0,856	0,939	0,999	0,111	0,922	0,198	1,120
556	8102	57685	348474	459	0,858	0,946	0,999	0,123	0,923	0,218	1,141
557	11259	54528	348175	758	0,865	0,937	0,998	0,171	0,926	0,289	1,216
558	9593	56194	348595	338	0,861	0,966	0,999	0,146	0,925	0,253	1,178
565	10806	54981	348389	544	0,864	0,952	0,998	0,164	0,926	0,280	1,206
566	18992	46795	347570	1363	0,881	0,933	0,996	0,289	0,935	0,441	1,376
567	19724	46063	347338	1595	0,883	0,925	0,995	0,300	0,936	0,453	1,389
568	26769	39018	339579	9354	0,897	0,741	0,973	0,407	0,934	0,525	1,459
655	5929	59858	348503	430	0,853	0,932	0,999	0,090	0,920	0,164	1,085
656	6398	59389	348438	495	0,854	0,928	0,999	0,097	0,921	0,176	1,097
657	7310	58477	348339	594	0,856	0,925	0,998	0,111	0,922	0,198	1,120
658	7782	58005	348316	617	0,857	0,927	0,998	0,118	0,922	0,210	1,132
665	14634	51153	347394	1539	0,872	0,905	0,996	0,222	0,930	0,357	1,287
666	14968	50819	347247	1686	0,872	0,899	0,995	0,228	0,930	0,363	1,293
667	17412	48375	347107	1826	0,878	0,905	0,995	0,265	0,933	0,410	1,342
668	18750	47037	346809	2124	0,881	0,898	0,994	0,285	0,934	0,433	1,367
675	21344	44443	344748	4185	0,886	0,836	0,988	0,324	0,934	0,467	1,402
676	21554	44233	343294	5639	0,886	0,793	0,984	0,328	0,932	0,464	1,396
677	21585	44202	344373	4560	0,886	0,826	0,987	0,328	0,934	0,470	1,403
678	22463	43324	343220	5713	0,888	0,797	0,984	0,341	0,933	0,478	1,411
685	25266	40521	340581	8352	0,894	0,752	0,976	0,384	0,933	0,508	1,441
686	20765	45022	347232	1701	0,885	0,924	0,995	0,316	0,937	0,471	1,408
687	26742	39045	338883	10050	0,897	0,727	0,971	0,406	0,932	0,521	1,454
688	27176	38611	337538	11395	0,897	0,705	0,967	0,413	0,931	0,521	1,452

SILLA CON MEJORAS											
BW	TP	FN	TN	FP	P0	P1	R0	R1	FS0	FS1	SUMA
566	24736	22626	365097	2261	0,942	0,916	0,994	0,522	0,967	0,665	1,632
567	24674	22688	364449	2909	0,941	0,895	0,992	0,521	0,966	0,658	1,625
568	22539	24823	366563	795	0,937	0,966	0,998	0,476	0,966	0,638	1,604
576	24950	22412	361286	6072	0,942	0,804	0,983	0,527	0,962	0,637	1,599
577	25194	22168	360161	7197	0,942	0,778	0,980	0,532	0,961	0,632	1,593
578	25196	22144	359030	8350	0,942	0,751	0,977	0,532	0,959	0,623	1,582
586	27056	20306	355000	12358	0,946	0,686	0,966	0,571	0,956	0,624	1,580
587	27027	20335	353894	13464	0,946	0,667	0,963	0,571	0,954	0,615	1,570
588	27060	20302	353709	13649	0,946	0,665	0,963	0,571	0,954	0,615	1,569
666	21219	26143	366601	757	0,933	0,966	0,998	0,448	0,965	0,612	1,577
667	22123	24543	366152	1902	0,937	0,921	0,995	0,474	0,965	0,626	1,591
668	25612	22132	359982	6994	0,942	0,785	0,981	0,536	0,961	0,638	1,599
676	29285	20077	351364	13994	0,946	0,677	0,962	0,593	0,954	0,632	1,586
677	28765	21332	354123	10500	0,943	0,733	0,971	0,574	0,957	0,644	1,601
678	27872	20987	356612	9249	0,944	0,751	0,975	0,570	0,959	0,648	1,608
686	30760	16602	349763	17595	0,955	0,636	0,952	0,649	0,953	0,643	1,596
687	29887	15092	350012	19729	0,959	0,602	0,947	0,664	0,953	0,632	1,585
688	27782	17982	354421	14535	0,952	0,657	0,961	0,607	0,956	0,631	1,587

ANEXO 2

En este anexo se presentan unas tablas que contienen los resultados del análisis de los parámetros más significativos del método de *Mean Shift* utilizado en la integración de la información de movimiento en la segmentación.

Las tablas recogen los resultados obtenidos para cada una de las secuencias analizadas, y en base a los mismos se han seleccionado los parámetros utilizados en la integración final. Como se ha explicado a lo largo del proyecto, el valor de Suma más alto corresponde a la prueba realizada con mayor calidad.

La variable **num** que aparece en las tablas corresponde a los tres valores de los números de regiones establecidos en el algoritmo de integración. Por lo tanto el valor de num igual a 163 corresponde a un valor de minNumReg igual a 1, maxNumReg1 igual a 6 y maxNumReg igual a 3.

BAILE UNIMODAL CON MEJORAS											
num	TP	FN	TN	FP	P0	P1	R0	R1	FS0	FS1	SUMA
163	33552	20571	360128	469	0,946	0,986	0,999	0,620	0,972	0,761	1,733
164	37704	16419	359135	1462	0,956	0,963	0,996	0,697	0,976	0,808	1,784
165	39283	14840	357866	2731	0,960	0,935	0,992	0,726	0,976	0,817	1,793
166	39283	14840	357072	3525	0,960	0,918	0,990	0,726	0,975	0,811	1,785
173	34409	19714	360088	509	0,948	0,985	0,999	0,636	0,973	0,773	1,746
174	37658	16465	359100	1497	0,956	0,962	0,996	0,696	0,976	0,807	1,783
175	39070	15053	357934	2663	0,960	0,936	0,993	0,722	0,976	0,815	1,791
176	37953	16170	358593	2004	0,957	0,950	0,994	0,701	0,975	0,807	1,782
263	34824	19299	359891	706	0,949	0,980	0,998	0,643	0,973	0,777	1,750
264	37834	16289	359017	1580	0,957	0,960	0,996	0,699	0,976	0,809	1,785
265	38660	15463	358263	2334	0,959	0,943	0,994	0,714	0,976	0,813	1,789
266	40909	13214	357780	2817	0,964	0,936	0,992	0,756	0,978	0,836	1,814
273	37724	16399	359709	888	0,956	0,977	0,998	0,697	0,977	0,814	1,790
274	37885	16238	359020	1577	0,957	0,960	0,996	0,700	0,976	0,810	1,785
275	40845	13278	358278	2319	0,964	0,946	0,994	0,755	0,979	0,840	1,818
276	40855	13268	357803	2794	0,964	0,936	0,992	0,755	0,978	0,836	1,814
283	39882	14241	359140	1457	0,962	0,965	0,996	0,737	0,979	0,836	1,814
284	38199	15924	358632	1965	0,957	0,951	0,995	0,706	0,976	0,810	1,786
285	41223	12900	358204	2393	0,965	0,945	0,993	0,762	0,979	0,844	1,823
286	41814	12309	357505	3092	0,967	0,931	0,991	0,773	0,979	0,844	1,823
363	39830	14293	358166	2431	0,962	0,942	0,993	0,736	0,977	0,826	1,804
364	38000	16123	357953	2644	0,957	0,935	0,993	0,702	0,974	0,802	1,776
365	39022	15101	357372	3225	0,959	0,924	0,991	0,721	0,975	0,810	1,785
366	39033	15090	356678	3919	0,959	0,909	0,989	0,721	0,974	0,804	1,778
373	38129	15994	357689	2908	0,957	0,929	0,992	0,704	0,974	0,801	1,776
374	38121	16002	357941	2656	0,957	0,935	0,993	0,704	0,975	0,803	1,778
375	39188	14935	357327	3270	0,960	0,923	0,991	0,724	0,975	0,812	1,787
376	41065	13058	356904	3693	0,965	0,917	0,990	0,759	0,977	0,831	1,808
383	38178	15945	357662	2935	0,957	0,929	0,992	0,705	0,974	0,802	1,776
384	38282	15841	357641	2956	0,958	0,928	0,992	0,707	0,974	0,803	1,777
385	41344	12779	357244	3353	0,965	0,925	0,991	0,764	0,978	0,837	1,815
386	39209	14914	356940	3657	0,960	0,915	0,990	0,724	0,975	0,809	1,783
463	39877	14246	357876	2721	0,962	0,936	0,992	0,737	0,977	0,825	1,801
464	40387	13736	357121	3476	0,963	0,921	0,990	0,746	0,976	0,824	1,801
465	40939	13187	357031	3563	0,964	0,920	0,990	0,756	0,977	0,830	1,807
466	41025	13098	356661	3936	0,965	0,912	0,989	0,758	0,977	0,828	1,805
473	39865	14258	357876	2721	0,962	0,936	0,992	0,737	0,977	0,824	1,801
474	40015	14108	357222	3375	0,962	0,922	0,991	0,739	0,976	0,821	1,797
475	40801	13322	356997	3600	0,964	0,919	0,990	0,754	0,977	0,828	1,805
476	41311	12812	356343	4254	0,965	0,907	0,988	0,763	0,977	0,829	1,805
483	40220	13903	357579	3018	0,963	0,930	0,992	0,743	0,977	0,826	1,803
484	38307	15816	358988	1609	0,958	0,960	0,996	0,708	0,976	0,815	1,791

BAILE CON MEJORAS											
num	TP	FN	TN	FP	P0	P1	R0	R1	FS0	FS1	SUMA
164	31359	19886	362457	1018	0,948	0,969	0,997	0,612	0,972	0,750	1,722
165	32541	18704	362065	1410	0,951	0,958	0,996	0,635	0,973	0,764	1,737
166	31795	19450	361452	2023	0,949	0,940	0,994	0,620	0,971	0,748	1,719
173	30198	21047	362832	643	0,945	0,979	0,998	0,589	0,971	0,736	1,707
174	33162	18083	362326	1149	0,952	0,967	0,997	0,647	0,974	0,775	1,749
175	33030	18215	361408	2067	0,952	0,941	0,994	0,645	0,973	0,765	1,738
176	33401	17844	360987	2488	0,953	0,931	0,993	0,652	0,973	0,767	1,739
183	28739	22506	362918	557	0,942	0,981	0,998	0,561	0,969	0,714	1,683
184	31484	19761	362279	1196	0,948	0,963	0,997	0,614	0,972	0,750	1,722
185	33127	18118	361595	1880	0,952	0,946	0,995	0,646	0,973	0,768	1,741
186	33518	17727	360784	2691	0,953	0,926	0,993	0,654	0,972	0,767	1,739
193	31836	19409	362804	671	0,949	0,979	0,998	0,621	0,973	0,760	1,733
194	33200	18045	362236	1239	0,953	0,964	0,997	0,648	0,974	0,775	1,749
195	33440	17805	361278	2197	0,953	0,938	0,994	0,653	0,973	0,770	1,743
196	33517	17728	360733	2742	0,953	0,924	0,992	0,654	0,972	0,766	1,738
264	25698	25265	354763	8994	0,934	0,741	0,975	0,504	0,954	0,600	1,554
265	31679	19566	361698	1777	0,949	0,947	0,995	0,618	0,971	0,748	1,719
266	33749	17496	361185	2290	0,954	0,936	0,994	0,659	0,973	0,773	1,747
274	33327	17918	362334	1141	0,953	0,967	0,997	0,650	0,974	0,778	1,752
275	33359	17886	361565	1910	0,953	0,946	0,995	0,651	0,973	0,771	1,745
276	33554	17691	360856	2619	0,953	0,928	0,993	0,655	0,973	0,768	1,740
284	33057	18188	362383	1092	0,952	0,968	0,997	0,645	0,974	0,774	1,748
285	33410	17835	361485	1990	0,953	0,944	0,995	0,652	0,973	0,771	1,745
286	33791	17454	361034	2441	0,954	0,933	0,993	0,659	0,973	0,773	1,746
294	33438	17807	362221	1254	0,953	0,964	0,997	0,653	0,974	0,778	1,753
295	33553	17692	361044	2431	0,953	0,932	0,993	0,655	0,973	0,769	1,742
296	33734	17511	360781	2694	0,954	0,926	0,993	0,658	0,973	0,770	1,742
364	33376	17869	362292	1183	0,953	0,966	0,997	0,651	0,974	0,778	1,752
365	31703	19542	361765	1710	0,949	0,949	0,995	0,619	0,971	0,749	1,720
366	31949	19296	361410	2065	0,949	0,939	0,994	0,623	0,971	0,749	1,721
374	31658	19587	362517	958	0,949	0,971	0,997	0,618	0,972	0,755	1,727
375	33497	17748	361092	2383	0,953	0,934	0,993	0,654	0,973	0,769	1,742
376	33795	17450	360920	2555	0,954	0,930	0,993	0,659	0,973	0,772	1,745
384	32038	19207	362127	1348	0,950	0,960	0,996	0,625	0,972	0,757	1,730
385	33501	17744	360834	2641	0,953	0,927	0,993	0,654	0,973	0,767	1,739
386	33584	17661	360402	3073	0,953	0,916	0,992	0,655	0,972	0,764	1,736
394	31982	19263	362092	1383	0,949	0,959	0,996	0,624	0,972	0,756	1,728
395	33465	17780	361476	1999	0,953	0,944	0,995	0,653	0,973	0,772	1,745
396	33540	17705	360480	2995	0,953	0,918	0,992	0,655	0,972	0,764	1,736
464	31939	19306	362280	1195	0,949	0,964	0,997	0,623	0,972	0,757	1,730
465	31690	19555	361804	1671	0,949	0,950	0,995	0,618	0,972	0,749	1,721
466	31778	19467	361293	2182	0,949	0,936	0,994	0,620	0,971	0,746	1,717
474	33687	17558	361825	1650	0,954	0,953	0,995	0,657	0,974	0,778	1,752
475	33698	17547	361203	2272	0,954	0,937	0,994	0,658	0,973	0,773	1,746
476	33765	17480	361024	2451	0,954	0,932	0,993	0,659	0,973	0,772	1,745
484	33475	17770	361741	1734	0,953	0,951	0,995	0,653	0,974	0,774	1,748
485	33489	17756	361206	2269	0,953	0,937	0,994	0,654	0,973	0,770	1,743
486	33625	17620	360485	2990	0,953	0,918	0,992	0,656	0,972	0,765	1,738

TENIS CON MEJORAS											
num	TP	FN	TN	FP	P0	P1	R0	R1	FS0	FS1	SUMA
163	7022	24701	381849	1148	0,939	0,859	0,997	0,221	0,967	0,352	1,319
164	9086	22637	381809	1188	0,944	0,884	0,997	0,286	0,970	0,433	1,402
165	9144	22579	380715	2282	0,944	0,800	0,994	0,288	0,968	0,424	1,392
166	9691	22032	379500	3497	0,945	0,735	0,991	0,305	0,967	0,432	1,399
173	9368	22355	381761	1236	0,945	0,883	0,997	0,295	0,970	0,443	1,413
174	9068	22655	381726	1271	0,944	0,877	0,997	0,286	0,970	0,431	1,401
175	9516	22207	380432	2565	0,945	0,788	0,993	0,300	0,968	0,434	1,403
176	10691	21032	379354	3643	0,947	0,746	0,990	0,337	0,969	0,464	1,433
183	8042	23681	381746	1251	0,942	0,865	0,997	0,254	0,968	0,392	1,361
184	9145	22578	381679	1318	0,944	0,874	0,997	0,288	0,970	0,434	1,403
185	9224	22499	381058	1939	0,944	0,826	0,995	0,291	0,969	0,430	1,399
186	10727	20996	379351	3646	0,948	0,746	0,990	0,338	0,969	0,465	1,434
193	7929	23794	381738	1259	0,941	0,863	0,997	0,250	0,968	0,388	1,356
194	9110	22613	381626	1371	0,944	0,869	0,996	0,287	0,970	0,432	1,401
195	9703	22020	380803	2194	0,945	0,816	0,994	0,306	0,969	0,445	1,414
196	11804	19919	378867	4130	0,950	0,741	0,989	0,372	0,969	0,495	1,465
263	5616	26107	381854	1143	0,936	0,831	0,997	0,177	0,966	0,292	1,257
264	9335	22388	381787	1210	0,945	0,885	0,997	0,294	0,970	0,442	1,412
265	9359	22364	381276	1721	0,945	0,845	0,996	0,295	0,969	0,437	1,407
266	11059	20664	379519	3478	0,948	0,761	0,991	0,349	0,969	0,478	1,447
273	9220	22503	381774	1223	0,944	0,883	0,997	0,291	0,970	0,437	1,407
274	9283	22440	381678	1319	0,944	0,876	0,997	0,293	0,970	0,439	1,408
275	9461	22262	381023	1974	0,945	0,827	0,995	0,298	0,969	0,438	1,408
276	10847	20876	378799	4198	0,948	0,721	0,989	0,342	0,968	0,464	1,432
283	9232	22491	381742	1255	0,944	0,880	0,997	0,291	0,970	0,437	1,407
284	9448	22275	381584	1413	0,945	0,870	0,996	0,298	0,970	0,444	1,414
285	9533	22190	381150	1847	0,945	0,838	0,995	0,301	0,969	0,442	1,412
286	9458	22158	381985	1119	0,945	0,894	0,997	0,299	0,970	0,448	1,419
383	9214	21548	381142	2816	0,946	0,766	0,993	0,300	0,969	0,431	1,400
384	9125	22056	380253	3286	0,945	0,735	0,991	0,293	0,968	0,419	1,386
385	9145	22415	380541	2619	0,944	0,777	0,993	0,290	0,968	0,422	1,390
386	9136	21549	381256	2779	0,947	0,767	0,993	0,298	0,969	0,429	1,398
463	9025	21458	381658	2579	0,947	0,778	0,993	0,296	0,969	0,429	1,398
464	9325	22563	381547	1285	0,944	0,879	0,997	0,292	0,970	0,439	1,409
465	9235	21458	381224	2803	0,947	0,767	0,993	0,301	0,969	0,432	1,401
466	9214	22145	380456	2905	0,945	0,760	0,992	0,294	0,968	0,424	1,392
473	9006	22121	380141	3452	0,945	0,723	0,991	0,289	0,967	0,413	1,381
474	9145	21565	381142	2868	0,946	0,761	0,993	0,298	0,969	0,428	1,397
475	9356	20104	380995	4265	0,950	0,687	0,989	0,318	0,969	0,434	1,403
476	9124	21411	380110	4075	0,947	0,691	0,989	0,299	0,968	0,417	1,385
483	9245	22102	381412	1961	0,945	0,825	0,995	0,295	0,969	0,435	1,404
484	9124	21221	381475	2900	0,947	0,759	0,992	0,301	0,969	0,431	1,400

INGRAVIDEZ CON MEJORAS											
num	TP	FN	TN	FP	P0	P1	R0	R1	FS0	FS1	SUMA
163	14979	50808	348437	496	0,873	0,968	0,999	0,228	0,931	0,369	1,300
164	18107	47680	347553	1380	0,879	0,929	0,996	0,275	0,934	0,425	1,359
165	18681	47106	347108	1825	0,881	0,911	0,995	0,284	0,934	0,433	1,367
166	19270	46517	346916	2017	0,882	0,905	0,994	0,293	0,935	0,443	1,377
173	15307	50480	348223	710	0,873	0,956	0,998	0,233	0,932	0,374	1,306
174	18469	47318	347304	1629	0,880	0,919	0,995	0,281	0,934	0,430	1,364
175	19653	46134	347305	1628	0,883	0,923	0,995	0,299	0,936	0,451	1,387
176	21212	44474	346658	2376	0,886	0,899	0,993	0,323	0,937	0,475	1,412
183	17287	48500	347622	1311	0,878	0,930	0,996	0,263	0,933	0,410	1,343
184	18696	47091	347290	1643	0,881	0,919	0,995	0,284	0,934	0,434	1,369
185	18952	46835	346962	1971	0,881	0,906	0,994	0,288	0,934	0,437	1,371
186	21412	44375	345425	3508	0,886	0,859	0,990	0,325	0,935	0,472	1,407
193	15609	50178	348031	902	0,874	0,945	0,997	0,237	0,932	0,379	1,311
194	16519	49268	347204	1729	0,876	0,905	0,995	0,251	0,932	0,393	1,325
195	21470	44317	346429	2504	0,887	0,896	0,993	0,326	0,937	0,478	1,415
196	21581	44206	347132	1801	0,887	0,923	0,995	0,328	0,938	0,484	1,422
264	18365	47422	347616	1317	0,880	0,933	0,996	0,279	0,934	0,430	1,364
265	18441	47346	347183	1750	0,880	0,913	0,995	0,280	0,934	0,429	1,363
266	21298	44489	345617	3316	0,886	0,865	0,990	0,324	0,935	0,471	1,407
274	17326	48461	347398	1535	0,878	0,919	0,996	0,263	0,933	0,409	1,342
275	19612	46175	346997	1936	0,883	0,910	0,994	0,298	0,935	0,449	1,384
276	22124	43663	346659	2274	0,888	0,907	0,993	0,336	0,938	0,491	1,428
284	19391	46396	347281	1652	0,882	0,921	0,995	0,295	0,935	0,447	1,382
285	20267	45520	346992	1941	0,884	0,913	0,994	0,308	0,936	0,461	1,397
286	21465	44322	346772	2161	0,887	0,909	0,994	0,326	0,937	0,480	1,417
294	19036	46751	347230	1703	0,881	0,918	0,995	0,289	0,935	0,440	1,375
295	20216	45571	346475	2458	0,884	0,892	0,993	0,307	0,935	0,457	1,392
296	20897	44890	346415	2518	0,885	0,892	0,993	0,318	0,936	0,469	1,404
364	18328	47459	347361	1572	0,880	0,921	0,995	0,279	0,934	0,428	1,362
365	21201	44586	347043	1890	0,886	0,918	0,995	0,322	0,937	0,477	1,414
366	19849	45938	346759	2174	0,883	0,901	0,994	0,302	0,935	0,452	1,387
374	18307	47480	347281	1652	0,880	0,917	0,995	0,278	0,934	0,427	1,361
375	19031	46756	346927	2006	0,881	0,905	0,994	0,289	0,934	0,438	1,373
376	21501	44286	346548	2385	0,887	0,900	0,993	0,327	0,937	0,480	1,416
384	18454	47333	347196	1737	0,880	0,914	0,995	0,281	0,934	0,429	1,363
385	20372	45415	346817	2116	0,884	0,906	0,994	0,310	0,936	0,462	1,397
386	19578	46209	347822	1111	0,883	0,946	0,997	0,298	0,936	0,453	1,389

SILLA CON MEJORAS											
num	TP	FN	TN	FP	P0	P1	R0	R1	FS0	FS1	SUMA
163	24093	23269	365831	1527	0,940	0,940	0,996	0,509	0,967	0,660	1,627
164	24747	22615	365627	1731	0,942	0,935	0,995	0,523	0,968	0,670	1,638
165	24844	22518	362670	4688	0,942	0,841	0,987	0,525	0,964	0,646	1,610
166	24851	22511	363923	3435	0,942	0,879	0,991	0,525	0,966	0,657	1,623
263	19687	27675	366735	623	0,930	0,969	0,998	0,416	0,963	0,582	1,545
264	22830	24532	366682	676	0,937	0,971	0,998	0,482	0,967	0,644	1,611
265	22490	24872	365772	1586	0,936	0,934	0,996	0,475	0,965	0,630	1,595
266	25425	21938	362324	5033	0,943	0,835	0,986	0,537	0,964	0,653	1,618
273	19745	27617	366663	695	0,930	0,966	0,998	0,417	0,963	0,582	1,545
274	19970	27392	366653	705	0,930	0,966	0,998	0,422	0,963	0,587	1,550
275	24704	22658	364436	2922	0,941	0,894	0,992	0,522	0,966	0,659	1,625
276	25406	21956	363392	3966	0,943	0,865	0,989	0,536	0,966	0,662	1,628
283	19704	27658	366660	698	0,930	0,966	0,998	0,416	0,963	0,582	1,544
284	24102	23260	366631	727	0,940	0,971	0,998	0,509	0,968	0,668	1,636
285	24597	22765	366054	1304	0,941	0,950	0,996	0,519	0,968	0,671	1,640
286	24971	22391	362084	5274	0,942	0,826	0,986	0,527	0,963	0,644	1,607
293	19748	27614	366878	480	0,930	0,976	0,999	0,417	0,963	0,584	1,547
294	24450	22912	365821	1537	0,941	0,941	0,996	0,516	0,968	0,667	1,634
295	22992	24370	363776	3582	0,937	0,865	0,990	0,485	0,963	0,622	1,585
296	24873	22489	360122	7236	0,941	0,775	0,980	0,525	0,960	0,626	1,586
363	19333	28029	366664	694	0,929	0,965	0,998	0,408	0,962	0,574	1,536
364	19958	27404	366658	700	0,930	0,966	0,998	0,421	0,963	0,587	1,550
365	24582	22780	364288	3070	0,941	0,889	0,992	0,519	0,966	0,655	1,621
366	24972	22390	362355	5003	0,942	0,833	0,986	0,527	0,964	0,646	1,609
373	19541	27821	366953	405	0,930	0,980	0,999	0,413	0,963	0,581	1,544
374	22865	24497	366553	805	0,937	0,966	0,998	0,483	0,967	0,644	1,610
375	23103	24259	364203	3155	0,938	0,880	0,991	0,488	0,964	0,628	1,591
376	24974	22388	362261	5097	0,942	0,831	0,986	0,527	0,963	0,645	1,608
383	22472	24890	366622	736	0,936	0,968	0,998	0,474	0,966	0,637	1,603
384	22825	24537	366633	725	0,937	0,969	0,998	0,482	0,967	0,644	1,610
385	24586	22776	365339	2019	0,941	0,924	0,995	0,519	0,967	0,665	1,632
386	24973	22389	361353	6005	0,942	0,806	0,984	0,527	0,962	0,638	1,600
393	22395	24967	366590	768	0,936	0,967	0,998	0,473	0,966	0,635	1,601
394	22766	24596	365745	1613	0,937	0,934	0,996	0,481	0,965	0,635	1,600
395	24661	22701	363221	4137	0,941	0,856	0,989	0,521	0,964	0,648	1,612
396	24872	22490	362037	5321	0,942	0,824	0,986	0,525	0,963	0,641	1,604
463	22694	24668	365353	2005	0,937	0,919	0,995	0,479	0,965	0,630	1,595
464	24779	22583	364782	2576	0,942	0,906	0,993	0,523	0,967	0,663	1,630
465	24610	22752	365356	2002	0,941	0,925	0,995	0,520	0,967	0,665	1,633

PRESUPUESTO

1) Ejecución Material

- Compra de ordenador personal (Software incluido) 2.000 €
- Material de oficina 150 €
- Total de ejecución material 2.150 €

2) Gastos generales

- 16 % sobre Ejecución Material 352 €

3) Beneficio Industrial

- 6 % sobre Ejecución Material 132 €

4) Honorarios Proyecto

- 700 horas a 15 € / hora 10.500 €

5) Material fungible

- Gastos de impresión 200 €
- Encuadernación 20 €

6) Subtotal del presupuesto

- Subtotal Presupuesto 13.354 €

7) I.V.A. aplicable

- 16% Subtotal Presupuesto 2136.64 €

8) Total presupuesto

- Total Presupuesto 15.490,64 €

Madrid, Julio de 2010

El Ingeniero Jefe de Proyecto

Fdo.: Pablo Cervera Benito
Ingeniero Superior de Telecomunicación

PLIEGO DE CONDICIONES

Este documento contiene las condiciones legales que guiarán la realización, en este proyecto, de análisis comparativo de técnicas de segmentación de secuencias de video basadas en modelado de fondo 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. Anexos

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.