



Universidad  
Carlos III de Madrid

Departamento de Teoría de la Señal y Comunicaciones

TRABAJO DE FIN DE GRADO

# Detección de Saliencia en Secuencias de Vídeo Basado en Segmentación

Autor: Diego Aramendi Campón

Tutor: Carmen Peláez Moreno

Leganés, junio de 2014



Título: Grado en Ingeniería de Sistemas Audiovisuales  
Autor: Diego Aramendi Campón  
Tutor: Carmen Peláez Moreno

## EL TRIBUNAL

Presidente: Marcelo Gabriel Bagnulo Braun

Vocal: Victor Pedro Gil Jimenez

Secretario: Mario García Valderas

Realizado el acto de defensa y lectura del Proyecto Fin de Carrera el día 11 de Julio de 2014 en Leganés, en la Escuela Politécnica Superior de la Universidad Carlos III de Madrid, acuerda otorgarle la CALIFICACIÓN de

VOCAL

SECRETARIO

PRESIDENTE



# Agradecimientos

*Agradezco a mi tutora Carmen Peláez del departamento de Teoría de la Señal y Comunicaciones la ayuda prestada para poder llevar a cabo este proyecto, así como los puntos de vista de Carlos Ruiz.*

*Agradezco también a mi familia la confianza que han depositado en mí durante estos años estudiando y en especial el apoyo de mi madre en los momentos complicados a lo largo de todo este tiempo.*



# Resumen

*Como dilucidó Platón, podemos distinguir dos mundos, el sensible y el inteligible. Al mundo sensible tenemos acceso a través de los sentidos y en él hay dos tipos de entidades, las imágenes de los objetos y los objetos físicos. Dentro de este mundo sensible, los seres humanos prestan mayor o menor atención a la información que les llega a través de los sentidos y es el contenido de esta información la que motiva al sujeto en sí a tenerlo en mayor o menor consideración y también despreciarlo.*

*En nuestro desarrollo, la idea que parte como innovadora es tratar de aplicar las técnicas de saliencia estática dentro de una imagen a una secuencia de vídeo.*

*En este proyecto se ha trabajado con la información visual, con imágenes y secuencias de imágenes para detectar sobre las mismas aquellas zonas que tendrían mayor relevancia para el sistema visual humano, estableciendo un modelo que aporta una mayor estabilidad en el cálculo de la saliencia para secuencias de vídeo, siendo ésta estabilidad el objetivo principal. Se ha llevado a cabo mediante diferentes técnicas de filtrado.*

*Mediante el cálculo de la saliencia en un vídeo se pretende obtener una salida que concuerde cualitativamente con la percepción de saliencia visual humana cuando se dispone a visualizar cualquier vídeo.*

**Palabras clave:** saliencia visual, segmentación, imagen, vídeo, detección, prominencia, percepción humana, sistema visual.





# Abstract

*As Platon elucidated, we are able to distinguish among two worlds, the sensitive world and the intelligible one. We can access to the sensitive world through our senses and in it there we able to find two entities, the objects' images and physical objects. Within the sensitive world, human beings pay more or less attention to the information they receive from their senses and the content of this information motivates the individual to either take it into consideration or despise it.*

*During this development, the idea which starts as innovative is to try to apply the static saliency techniques from an image to a video sequence.*

*Throughout this assumption it has been worked with visual information, also with images and sequences of images to detect those areas which would have more relevance to the human visual system establishing a model providing greater stability in the saliency computation for video sequences, being this the main goal. It was carried out using different filtering techniques.*

*The object of saliency computation in video is to achieve an output that agrees qualitatively with the perception of human visual saliency.*

**Keywords:** visual saliency, segmentation, image, video, detection, prominence, human perception, visual system.



# Índice general

<b>1. INTRODUCCIÓN Y OBJETIVOS .....</b>	<b>1</b>
1.1 Introducción .....	1
1.2 Introduction.....	3
1.3 Objetivos .....	4
1.4 Fases del desarrollo.....	4
1.5 Planificación .....	5
1.6 Medios empleados .....	7
1.7 Estructura de la memoria .....	7
<b>2. DETECCIÓN DE SALIENCIA .....</b>	<b>8</b>
2.1 Estado del arte.....	8
2.1.1 Saliencia .....	8
2.1.2 Segmentación.....	12
2.1.3 Tracking de objetos .....	15
<b>3. ANÁLISIS DE LA SALIENCIA OBTENIDA PARA DIFERENTES IMÁGENES CON     LA HERRAMIENTA DE MURATOV.....</b>	<b>17</b>
3.1 Introducción .....	17
3.2 Segmentación.....	18
3.3 Detección de saliencia .....	23

<b>4. ADAPTACIÓN DE LA HERRAMIENTA DE MURATOV PARA SECUENCIAS DE VÍDEO.....</b>	<b>26</b>
4.1 Introducción .....	26
4.2 Adaptación .....	27
4.3 Problemas encontrados .....	27
4.4 Propuestas, soluciones y alternativas de diseño .....	29
4.5 Resultados .....	44
<b>5. VIABILIDAD .....</b>	<b>45</b>
5.1 Entorno socio-económico .....	45
5.2 Marco regulador.....	47
5.3 Presupuesto .....	48
<b>6. CONCLUSIONES Y LÍNEAS FUTURAS.....</b>	<b>50</b>
6.1 Conclusiones .....	50
6.2 Conclusions.....	51
6.3 Líneas futuras.....	52
<b>7. BIBLIOGRAFÍA.....</b>	<b>54</b>
7.1 Referencias bibliográficas.....	54

# Índice de figuras

<b>Figura 1</b> Filosofías para la detección de saliencia Muratov O. <i>et al.</i> [5].....	2
<b>Figura 2</b> Arquitectura General del Modelo Itti <i>et al.</i> [2]. .....	9
<b>Figura 3</b> Detector de saliencia basado en segmentación, Muratov <i>et al.</i> [5]. .....	11
<b>Figura 4</b> Ejemplos de una buena segmentación según Bagon <i>et al.</i> [7]. ....	12
<b>Figura 5</b> Aprendizaje máquina para detectar caballos Levin, A. & Weiss, Y [18]. .....	13
<b>Figura 6</b> Coosegmentación de imágenes según Rother, C <i>et al.</i> [19].....	14
<b>Figura 7</b> Grafo de la segmentación de Tu, Z. <i>et al</i> [20].....	15
<b>Figura 8</b> Flujo óptico.....	16
<b>Figura 9</b> Segmentación variando el área mínima del segmento .....	20
<b>Figura 10</b> Modificación del Spatial bandwidth.....	22
<b>Figura 11</b> Ejemplos de Saliencia Muratov .....	24
<b>Figura 12</b> Imagen Vs Imagen Segmentada en planos consecutivos .....	30
<b>Figura 13</b> Variación en la saliencia para el mismo segmento.....	31

## ÍNDICE DE FIGURAS

<b>Figura 14</b> Cálculo de los centroides .....	32
<b>Figura 15</b> K-means sobre centroides.....	33
<b>Figura 16</b> Comparativa K-means con Frame .....	34
<b>Figura 17</b> Erosión .....	36
<b>Figura 18</b> Dilatación.....	37
<b>Figura 19</b> Filtrado temporal .....	39
<b>Figura 20</b> Comparación aplicando cierre sobre la segmentación de dos frames consecutivos.....	40
<b>Figura 21</b> Comparación de filtrado en la segmentación .....	41
<b>Figura 22</b> Comparación aplicando cierre sobre la segmentación de dos frames consecutivos.....	42
<b>Figura 23</b> Comparación filtrado en Saliencia .....	43
<b>Figura 24</b> Dónde resulta menos intrusiva la publicidad H. Liu <i>et al.</i> [21] ..	46

# Índice de tablas

<b>Tabla 1</b> Diagrama de Gantt Ideal .....	5
<b>Tabla 2</b> Diagrama de Gantt Real .....	6
<b>Tabla 3</b> Parámetros EDISON por Muratov .....	19
<b>Tabla 4</b> Parámetro modificado .....	20
<b>Tabla 5</b> Número de segmentos para el mismo frame variando el área .....	21
<b>Tabla 6</b> Comparativa de la modificación Spatial bandwidth .....	22
<b>Tabla 7</b> Parámetros fijos Muratov .....	23
<b>Tabla 8</b> Encuesta sobre saliencia.....	25
<b>Tabla 9</b> Correlación entre frames consecutivos .....	28
<b>Tabla 10</b> Resultados .....	44
<b>Tabla 11</b> Presupuesto.....	46





# Capítulo 1

## Introducción y objetivos

### 1.1 Introducción

La saliencia es un anglicismo proveniente de “saliency” para definir la condición o condiciones por las que un elemento es más notable o cobra una mayor importancia dentro de un conjunto.

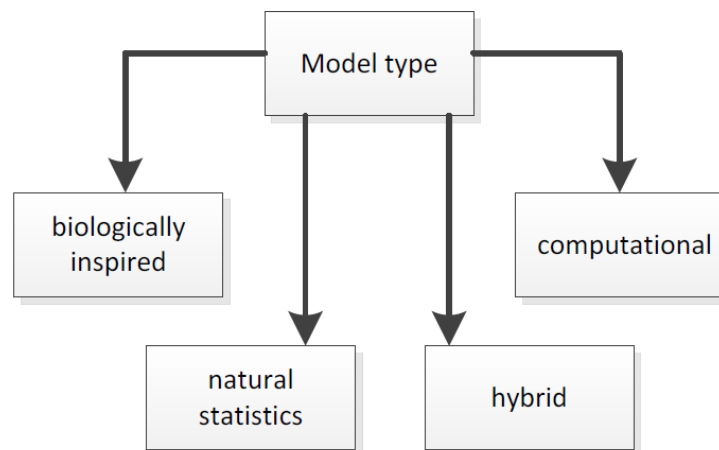
En psicología, la saliencia se utiliza en el estudio de la percepción y la cognición para referirse a cualquier estímulo, que por factores emocionales, motivacionales o cognitivos destaca sobre el resto y que no están necesariamente ligados a factores físicos como el tamaño. El hipocampo es la parte de nuestro cerebro que participa en la evaluación de la relevancia y el contexto, utilizando los recuerdos del pasado para filtrar los nuevos estímulos entrantes. La corteza entorrinal es la vía de entrada y salida al hipocampo ya que está muy ligada a los impulsos del ojo y el cerebro.

En este trabajo se aplica el término de saliencia a la percepción visual del ser humano. La capacidad del sistema visual humano en este aspecto es muy rápida, sin embargo, el modelado computacional de esta capacidad continúa siendo un problema a solucionar.

La capacidad de distinguir unos objetos de otros, su profundidad con respecto al observador, su movimiento, y otros, como el color, el tamaño o la forma, han motivado diferentes estudios de modelos psicológicos y técnicos.

Encontrar una solución matemática al problema de la saliencia es bastante improbable debido a la alta complejidad. Lograr un método intuitivo para determinar las regiones más interesantes de una imagen de una forma natural y lo más parecida al ojo humano es un prometedor enfoque para mejorar los sistemas de visión artificial.

Las aplicaciones de detección de saliencia se organizan a grandes rasgos en tres categorías, según Frintrop S. *et al.* [22]: la primera o de bajo nivel basadas en puntos o regiones de interés básicos, la categoría de nivel medio que utiliza el reconocimiento de objetos y una última categoría de alto nivel en la que se entrena el sistema con datos reales. Pero para cada aplicación existen diferentes enfoques a la hora de detectar la saliencia. Por un lado está el enfoque denominado como biológico que trata de analizar las características del ojo humano, cómo opera la atención y qué procesos suceden a la hora de observar. Otros enfoques son los basados en datos estadísticos de las imágenes como su histograma, color del pixel o el análisis de su espectro, mientras que otros analizan la información del conjunto computacionalmente. Por supuesto, también se ha trabajado combinando varios de estos enfoques para dar lugar a híbridos. Un esquema se muestra en la Figura 1.



**Figura 1 Filosofías para la detección de saliencia Muratov O. *et al.* [5]**

Dentro de los modelos podemos hablar de aquellos que calculan la saliencia de toda la imagen píxel a píxel y de modelos híbridos que

utilizan una segmentación previa de la imagen para posteriormente calcular su saliencia. En nuestro caso hemos trabajado con un modelo de saliencia híbrido basado en la segmentación que explicaremos más adelante.

## 1.2 Introduction

The saliency is a form to define the condition or conditions for an element that are more notable or receives a major importance inside a set of them.

In psychology the saliency is apply in the study of the perception and the cognition for referring to any stimulus that for emotional factors, motivational or cognitive emphasizes on the rest and that is not necessarily tied to physical factors as the size. The hippocampus is the part of our brain that takes part in the evaluation of the relevancy and the context, using the memory of the past to filter the new incoming stimulus. The entorhinal cortex is the route of entry and exit for the hippocampus since it is very tied to the impulses of the eye and the brain.

In this work saliency is applied to the visual perception of the human visual system. The visual human system capacity in this aspect is very fast, nevertheless, a computationally model of this capacity continues being a problem to solving.

The aptitude to distinguish a few objects from others, his depth over the observer, his movement, and others, as the color, the size or the form have motivated different studies of psychological and technical models.

To find a mathematical solution to the saliency problem is unlikely, due to the high complexity, to achieve an intuitive method to determine the most interesting regions of an image of a natural form and the most similar to the human eye it is a promising approach to improve the artificial vision systems.

The applications of saliency detection are organized in broad terms in three categories, according to Frintrop S. *et al.* [22]: first or low-level based in basic points or regions of interest, then an average-level that uses the recognition of objects and a last high-level category in which, the system is trained by actual data information. But different approaches exist for every application at the moment for detecting the saliency. On the one hand the approach is named biological, that tries to analyze the characteristics of the human eye, how it produces the attention and what processes happen at the

moment of observing. Other approaches are based on statistical information of the images as his histogram, the color of the pixel or the analysis of his spectrum, whereas others analyze the information of the set computationally. Off course also it has being worked by combining several of these approaches for a hybrid result. A schematic is shown in Figura 1.

Among the models, we can mention ones that compute the saliency on an image pixel by pixel and hybrid models which use a prior image segmentation for then calculate the saliency levels. In our case we have worked with a hybrid model of saliency-based segmentation which is explained later.

## 1.3 Objetivos

El objetivo principal del proyecto es, en primer lugar, identificar las dificultades asociadas a la extensión de métodos de saliencia estática a secuencias de vídeo y en segundo lugar, proponer y evaluar métodos para solucionarlos.

En nuestro desarrollo, la idea que parte como innovadora, es tratar de generar un modelo estable de detección de saliencia en secuencias vídeo a partir de un modelo de detección de saliencia estática basado en segmentación.

De este modo, obtener una salida estable a lo largo de dicha secuencia nos proporcionará una mejor concordancia cualitativa con la percepción de la saliencia por el sistema visual humano.

## 1.4 Fases del desarrollo

- Primera Fase. “Recopilación de bibliografía sobre detección de saliencia estática, percepción y estimación del movimiento y sobre eye-tracking”.
- Segunda Fase. “Puesta a punto del código disponible de la tesis doctoral de Muratov O. *et al.* [5]”.
- Tercera Fase. “Análisis de la saliencia obtenida para diferentes imágenes con la herramienta de Muratov”.

- Cuarta Fase. “Adaptación de la herramienta de Muratov para aplicarla en secuencias de vídeo y análisis de la saliencia a la salida”.
- Quinta Fase. “Aumentar la consistencia de la salida de saliencia a la herramienta modificada de Muratov”.
- Sexta Fase. “Documentación”.

## 1.5 Planificación

La planificación propuesta, en un principio fue la comentada en 1.4 hasta la quinta fase, añadiendo dos fases más, la fase 6 en la que se pretendía analizar la estimación de movimiento de las secuencias de vídeo y la fase 7 en la que se pretendía entrenar el modelo. La documentación de la memoria pasa a la fase 8 como se ve en la Tabla 1:

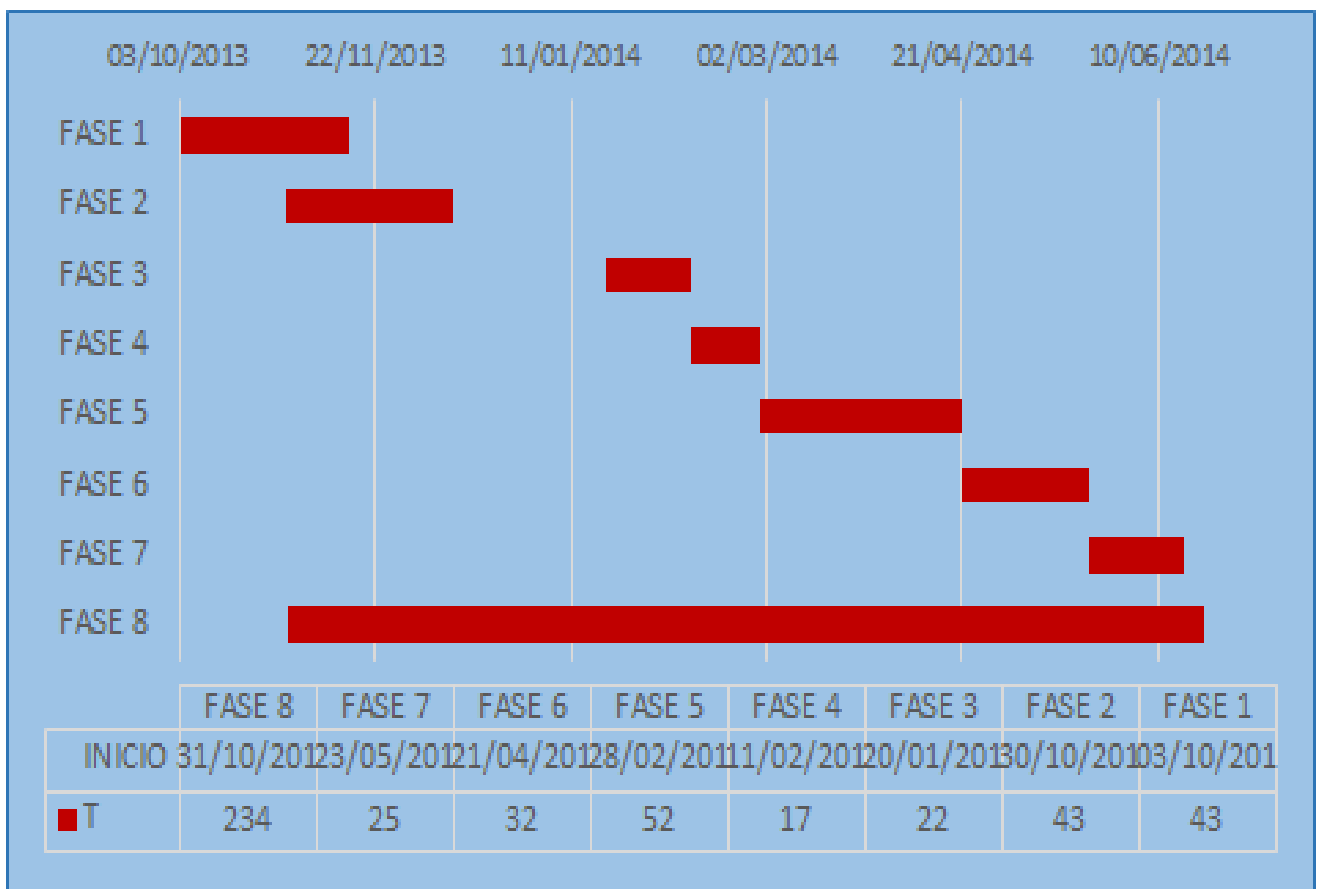
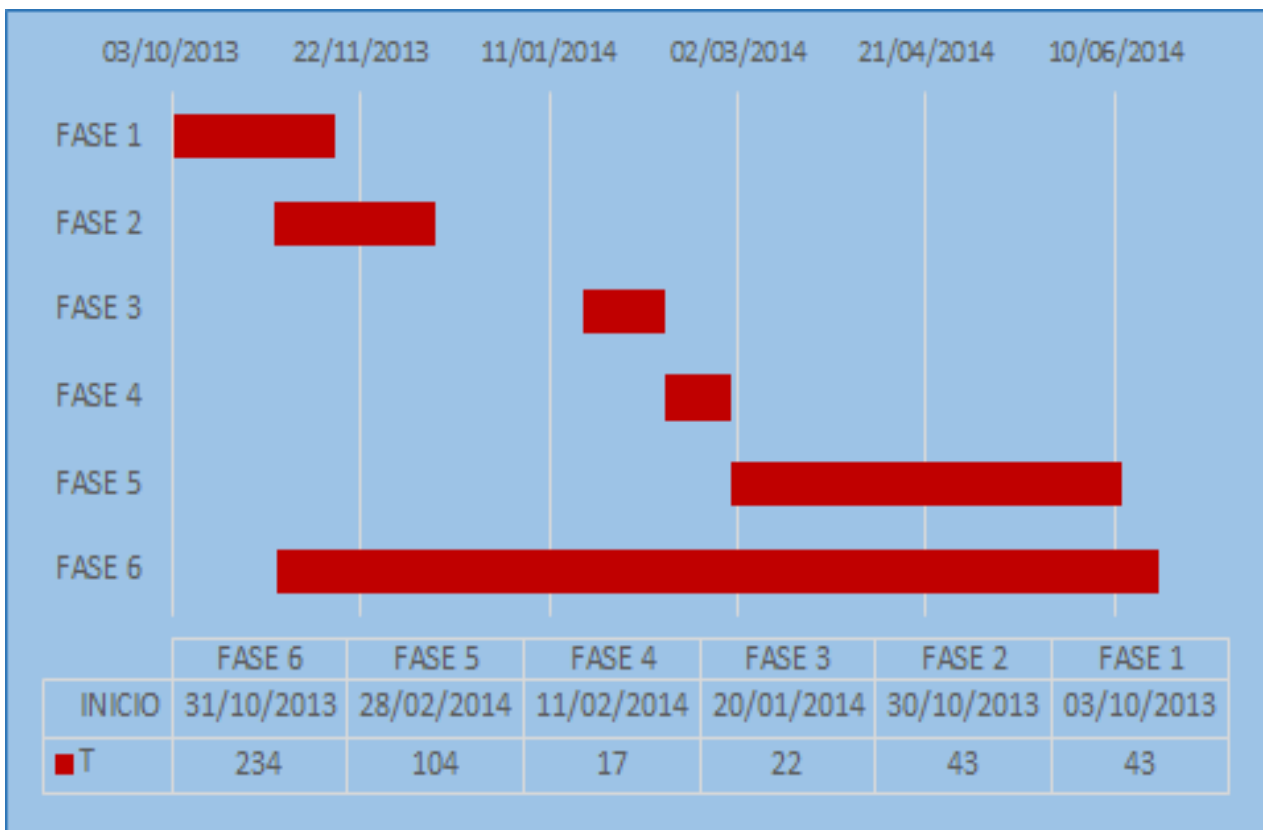


Tabla 1 Diagrama de Gantt Ideal

Desde el primer momento nos habría gustado llegar a aplicar las técnicas de saliencia estática a un vídeo y combinarlo con las técnicas de saliencia basadas en la estimación de movimiento y el flujo óptico, para finalmente entrenar el algoritmo con datos reales de eye-tracking, pero no sabíamos los desafíos que podríamos encontrarnos en el camino. Esto nos ha supuesto diversos problemas que se han explicado a lo largo de esta memoria, por los cuales no se han logrado todos los objetivos que pretendíamos en un principio.

Se han realizado reuniones semanales con el tutor para intentar la consecución de estos objetivos.

Las fases finalmente llevadas a cabo y que fueron expuestas en el apartado 1.4 se reflejan en el diagrama de la Tabla 2 con su duración:



**Tabla 2 Diagrama de Gantt Real**

## 1.6 Medios empleados

Se ha empleado el código de la tesis doctoral de Muratov O. *et al.* [5] de la Universidad de Trento.

Se ha utilizado la tecnología de Matlab para todo el proyecto y algunas operaciones utilizan una compilación en lenguaje C++.

No ha sido necesario obtener por nosotros mismos los datos de vídeos de prueba, puesto que se nos han facilitado también desde la Universidad de Trento.

## 1.7 Estructura de la memoria

El Capítulo 2 versa sobre la detección de saliencia. Se lleva a cabo un análisis comparativo del estado del arte para entrar en la escena de las investigaciones más relevantes llevadas a cabo desde el comienzo del estudio y el análisis de la saliencia hasta los avances más actuales, siempre en el término que nos interesa a nosotros, que es la imagen digital.

El Capítulo 3 explica el algoritmo de saliencia de Muratov [5] así como las herramientas que utiliza para su desarrollo y los parámetros de libre configuración de que dispone.

El Capítulo 4 explica cómo se ha realizado la adaptación de la herramienta de saliencia estática para que funcionase en secuencias de vídeo, los problemas encontrados y las soluciones propuestas, tanto las que han tenido un buen resultado como las que no, además de plantear alternativas de diseño de las soluciones.

El Capítulo 5 habla sobre la viabilidad del proyecto, la importancia de las aplicaciones basadas en saliencia en diferentes ámbitos y cómo y dónde se han venido utilizando estas aplicaciones de detección

El Capítulo 6 ofrece las conclusiones y las líneas futuras sobre el trabajo llevado a cabo.

Para terminar la memoria se incluye toda la referencia bibliográfica que ha sido utilizada para este trabajo.

# Capítulo 2

## DetECCIÓN DE SALIENCIA

### 2.1 Estado del arte

#### 2.1.1 Saliencia

Los primeros estudios relacionados con la saliencia comienzan en la década de los ochenta, cuando se comienza a observar cómo los seres humanos perciben la velocidad en las secuencias de imágenes con movimientos y cómo les afecta, buscando modelos concordantes con la realidad de la percepción humana.

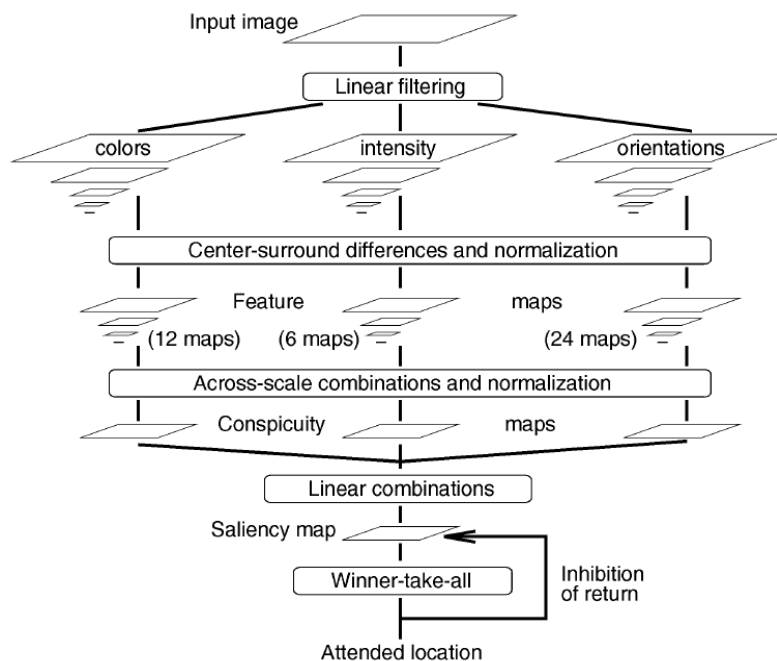
Un modelo subyacente que ha cobrado especial relevancia y sobre el que se sigue trabajando actualmente es el propuesto por Watson, A. B. & Ahumada, A. J. [1], que ha resultado uno de los pilares fundamentales en la investigación sobre la detección visual de movimiento, en el que se presenta un sensor de movimiento basado en algunas características de la percepción humana sobre el movimiento de objetos en una secuencia de imágenes y el análisis en frecuencia de las mismas.

Más de una década después encontramos el siguiente modelo revolucionario, pero esta vez los estudios se centran sobre la saliencia estática,



es decir, sin tener en cuenta el movimiento intentarán detectar qué zonas de la imagen cobran una mayor relevancia comparándolo cualitativamente con la percepción humana.

La idea y desarrollo principal es de Itti *et al.* [2], que publicó su investigación sobre la que se han basado otros muchos trabajos de visión artificial y neurobiología. El objetivo es predecir en qué puntos de la imagen el ser humano mostraría una mayor atención, es decir, trata de emular la actividad que lleva a cabo el córtex frontal humano. Se trata de un modelo jerárquico como el mostrado en la Figura 2 para imágenes estáticas. El mapa de saliencia se utiliza para codificar y combinar la información de cada punto saliente en la imagen.



**Figura 2 Arquitectura General del Modelo Itti et al. [2]**

La imagen se filtra y submuestra en diferentes niveles. Primero se realiza un filtrado lineal en bajo nivel basado en las características visuales que ofrecen la intensidad, la orientación y el color y posteriormente, se aplica lo que denomina como Center-Surround que a escala más pequeña calcula cada mapa de características para detectar elementos de diferentes tamaños. Este mapa se genera mediante pirámides gaussianas cuya varianza está en los márgenes de la escala en la que se trabaja. Por último se aplica una operación de normalización con un promedio de todos los máximos locales de cada mapa y finalmente se combinan linealmente para dar lugar a la imagen de saliencia sobre el que detectan con un algoritmo iterativo los focos más salientes.

Una vez más, tendremos que esperar unos cuantos años hasta la primera década del siglo XXI, que es cuando aparecen algunos desarrollos interesantes a partir del modelo de Itti *et al.* [2] y otras ideas completamente novedosas.

Otro modelo a destacar por relevancia y número de citas en artículos de investigación es el modelo de Harel *et al.* [3], que basándose en los modelos de Itti *et al.* [2], normalizando los mapas obtenidos, consiguen un mayor realce de aquellas zonas salientes al combinarlo con otros mapas de saliencia y ofrecer a la salida tanto el foco más saliente como aquellas zonas del foco con mayor relevancia de atención visual.

Un enfoque diferente y novedoso fue el propuesto por X. Hou and L. Zhang [4] que se basaron en el análisis espectral para generar un mapa de saliencia espacial.

En el último trienio la investigación se está desarrollando mucho más y se esperan noticias interesantes e innovadoras al respecto.

De los posibles modelos más actuales sobre detección de saliencia, se ha utilizado el modelo de 2013 propuesto por Muratov *et al.* [5], puesto que se disponía del código y derechos para su utilización. Este modelo propone dos métodos innovadores a la hora de detectar la saliencia: el primero está basado en los mapas de características visuales de cada objeto mediante una segmentación previa de los objetos de la imagen. El segundo, utiliza como herramienta principal la información de profundidad de cada uno de los objetos segmentados de la imagen. Combinando ambos métodos se obtiene como salida la imagen de saliencia.

Para Muratov *et al.* [5] es más importante extraer las características de cada objeto de interés de la imagen, que obtener las características de interés globales de la imagen píxel a píxel, porque de este modo logra reconocer objetos de interés dispersos dentro de la imagen, algo más parecido a lo que realizaría nuestro córtex. Por esta razón, hace uso de la extracción de objetos de la imagen a partir de un algoritmo de segmentación, aplicando primero la segmentación, posteriormente analizando la saliencia de cada segmento y logrando una estimación más precisa de la saliencia real para un observador humano. Se puede observar en el esquema presentado en la Figura 3.

El modelo de Muratov *et al.* [5] aplica y calcula varios parámetros. El primero de ellos es la importancia de cada color en la imagen, ya que atendiendo a estudios hace uso de los colores que son más relevantes para el ojo humano. Otro parámetro es meramente geométrico, se debe a que

normalmente se tiende a focalizar el objeto de interés en el centro de la imagen o del campo de visión, esto hace que sea una fuerte razón de peso, así como el tamaño del objeto, que normalmente hace que sea más saliente cuanto mayor tamaño tenga. El siguiente parámetro trabaja las oclusiones, es decir, aquellas regiones u objetos que ocuyen a otras cobran mayor relevancia saliente, puesto que se consideran como primer plano con respecto al resto que se tratan como fondo. Por último utiliza el método de detección de saliencia basado en la profundidad, en el que ésta se estima a través de la disposición espacial en 2D.

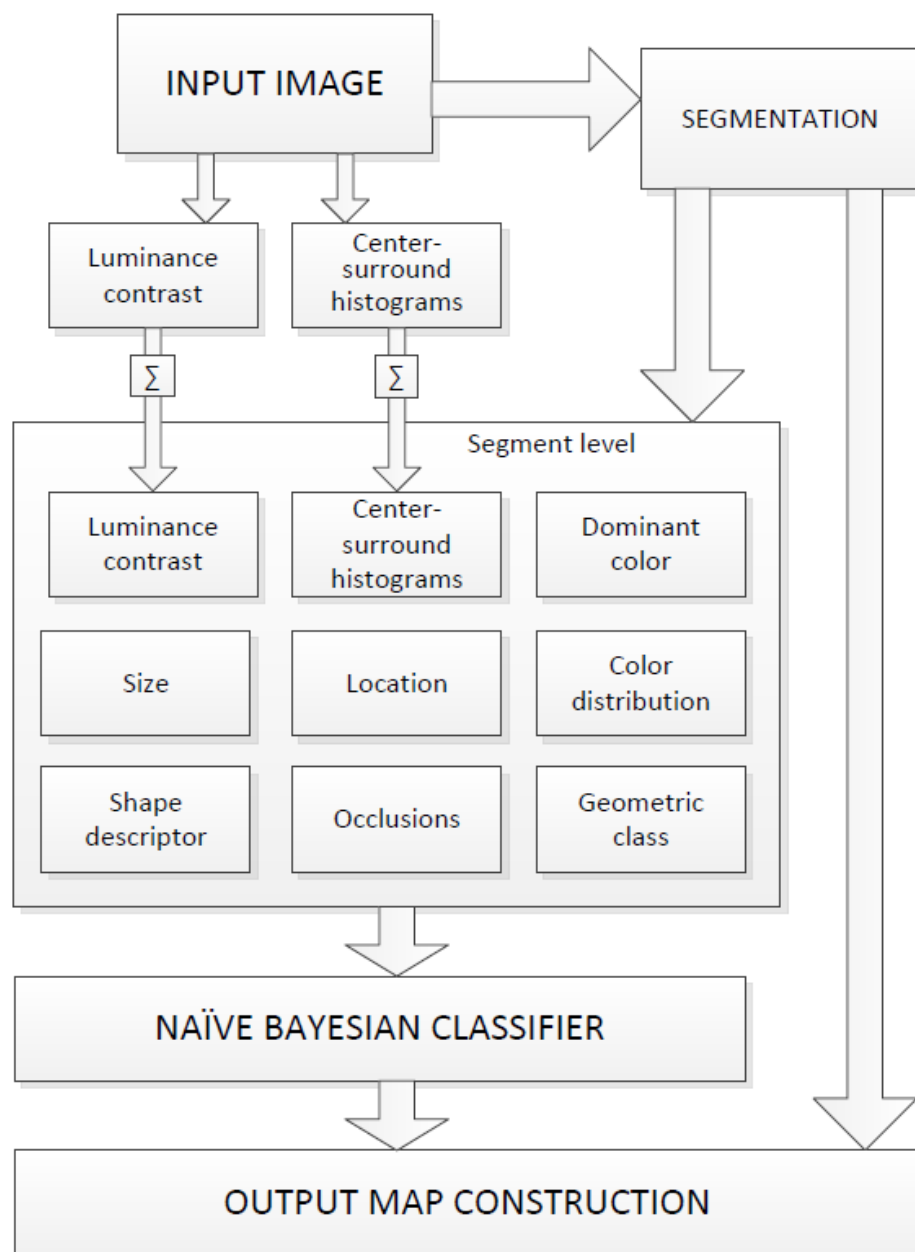


Figura 3 Detector de saliencia basado en segmentación, Muratov *et al.* [5]

## 2.1.2 Segmentación

Durante éste tiempo aparecen también las denominadas técnicas de segmentación, en las que se desarrollan numerosas herramientas que tratan de dividir una imagen por conjuntos de píxeles que guardan una determinada relación, por ejemplo pertenecer a un mismo objeto asignando una etiqueta a cada píxel de la imagen en función del segmento al que corresponden.

Las primeras técnicas en aparecer al respecto, son aquellas que separan el fondo de los objetos en primer plano grosso modo, generalmente umbralizando la imagen y generando una salida binaria con fondo o primer plano.

El modelo de Muratov *et al.* [5] utiliza la herramienta de segmentación avanzada EDISON, *edge detection and image segmentation system* [6], basada en la segmentación mean-shift, un algoritmo de clustering que sobre segmenta la imagen. A cada segmento se le asigna la mediana del desplazamiento medio de los píxeles que pertenecen a dicho segmento, posteriormente EDISON calcula los bordes de la imagen mediante el filtrado de Canny. Combinando ambas técnicas se obtiene a la salida la imagen segmentada. Esta herramienta contiene multitud de parámetros variables para el cálculo de la mean-shift y el cálculo de borde que modifican notablemente la segmentación global de la imagen.

Se han modelado numerosas técnicas para segmentar objetos, pero, ¿cómo sabemos si una segmentación es buena? Como indican Bagon *et al.* [7] depende del caso, una buena segmentación puede ser una región de color uniforme y en otro caso podría ser una región de textura similar, es decir, una composición de regiones similares pero independientes u otros objetos más complejos. Algunos ejemplos pueden observarse en la Figura 4.

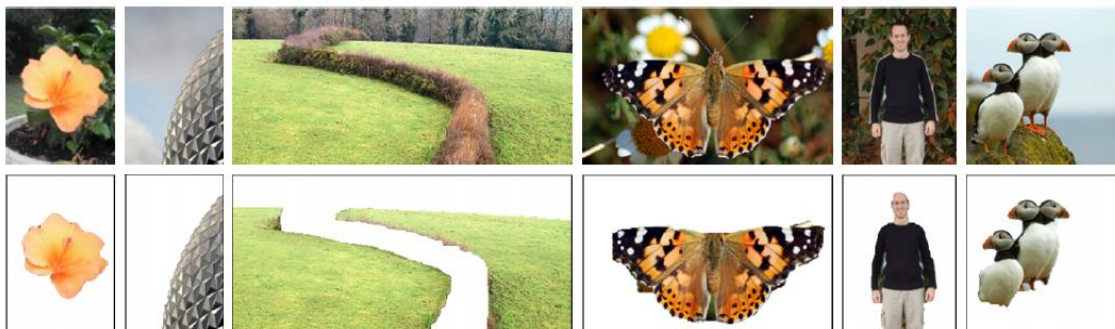
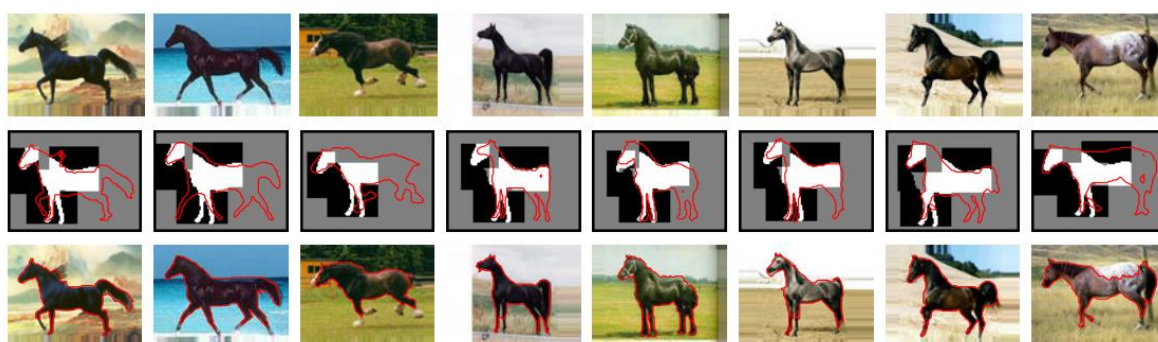


Figura 4 Ejemplos de una buena segmentación según Bagon *et al.* [7].

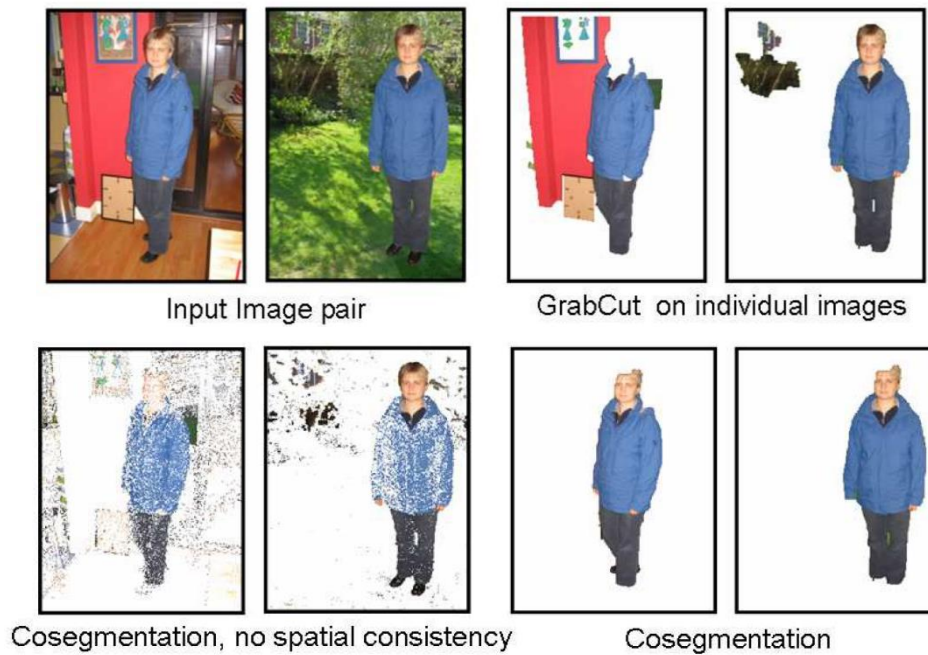
Esta diversidad de tipos de segmentos ha dado lugar a las diferentes técnicas de segmentación. Existen algoritmos que detectan regiones de colores uniformes (Shi, J. & Malik, J [8], Comaniciu, D. & Meer, P [9]), otros que extraen las características de textura (Malik, J *et al.* [10], Galum, M *et al.* [11]), y algoritmos para extraer segmentos de color basados en razonamientos empíricos (Kadir, T. & Brady, M. [12], Li, Y. *et al.* [13], Rother, C *et al.* [14]).

Hay otros que buscan simetrías en la imagen (Riklin-Raviv *et al.* [15]) y también los hay que se basan en alto nivel de semántica (aprendizaje máquina), es decir, mediante un conjunto de datos de entrenamiento son capaces de segmentar unos objetos muy determinados (Borestein, E. & Ullman, S. [16], Leibe, B & Schiele, B [17], Levin, A. & Weiss, Y [18] como se muestra en la Figura 5). Dentro de todos ellos están los parametrizados, que pueden ser controlados por el usuario o con una parametrización fija que impida su modificación.



**Figura 5 Aprendizaje máquina para detectar caballos Levin, A. & Weiss, Y [18]**

Otras variantes se han apoyado en el uso de varias imágenes para obtener la segmentación de un objeto. Se denomina cosegmentación al proceso en el que se lleva a cabo una segmentación simultanea de las partes comunes de ambas imágenes, Rother, C *et al.* [19] como puede observarse en la Figura 6.



**Figura 6 Coosegmentación de imágenes según Rother, C et al. [19]**

Como se ha explicado, hay un gran número de posibilidades a la hora de llevar a cabo una segmentación y no podemos considerarlo como un problema tecnológico resuelto. Puesto que esto es un elemento clave en el sistema de detección de saliencia que hemos adoptado, ha sido la causa de multitud de problemas a la hora de trasladar los modelos de saliencia estática a dinámica.

Por eso la ciencia necesita desarrollar un modelo que agrupe a todas ellas en una única herramienta, un modelo prometedor es el probabilístico propuesto por Tu, Z. *et al* [20] cuyo grafo se presenta en la Figura 7.

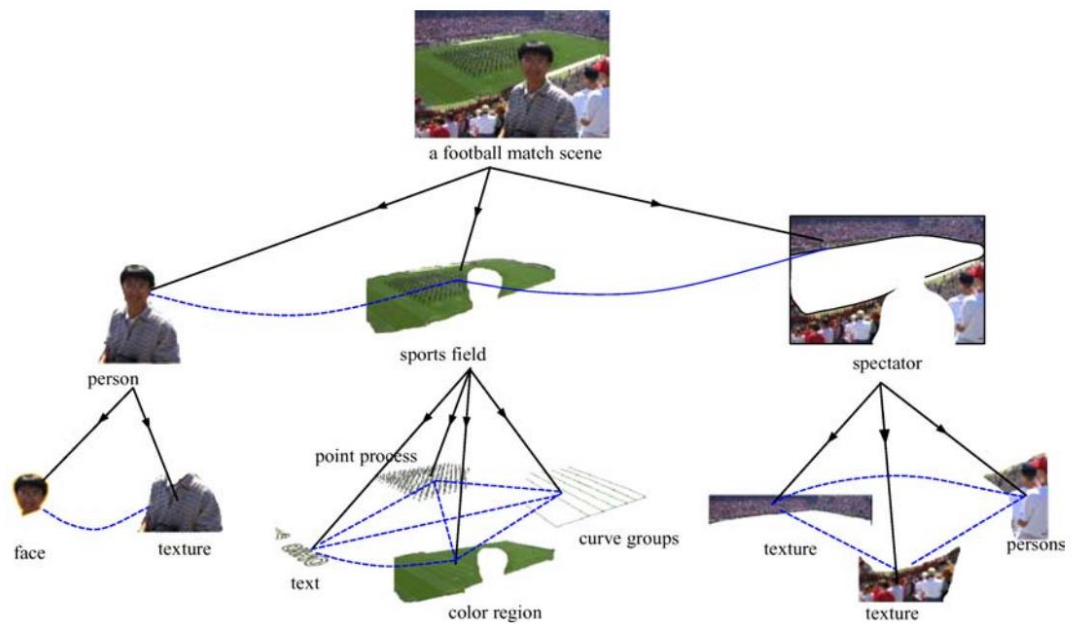


Figura 7 Grafo de la segmentación de Tu, Z. et al [20].

### 2.1.3 Tracking de objetos

La estimación de movimiento es un factor importantísimo cuando hablamos de vídeo, puesto que un vídeo es una secuencia de imágenes que representan escenas en movimiento, de no ser así pierde el significado.

Puesto que puede resultar de interés generar un descriptor de saliencia basado en el movimiento, hemos buscado sobre los detectores de tracking que mejor funcionan y a los que tenemos acceso al código libre. Entre todos debemos destacar el modelo de Liu [26].



**Figura 8 Flujo óptico**

En la Figura 8 se muestra un ejemplo de la detección de movimiento de Liu, donde el matiz indica la orientación, es decir, hacia dónde se mueve, y la saturación indica la magnitud del movimiento.

Se puede observar el vídeo completo para una secuencia de ejemplo en la siguiente dirección:

[\[https://www.youtube.com/watch?v=wrnjHQTjY\]](https://www.youtube.com/watch?v=wrnjHQTjY)



# Capítulo 3

## Análisis de la saliencia obtenida para diferentes imágenes con la herramienta de Muratov.

### 3.1 Introducción

La herramienta de Oleg Muratov está diseñada para obtener la saliencia de una imagen. Utiliza una segmentación paralela en la que se apoya para generar el mapa de saliencia y los descriptores. Posteriormente procederemos a explicarla.

En la tesis de Muratov [5], la solución más viable para la extracción de objetos a partir de imágenes es la segmentación de imágenes. En el caso de la detección de la saliencia tenemos 2 posibles formas de aplicar la segmentación:

- Por el cálculo de un mapa de saliencia y derivar un valor de la saliencia promedio de los segmentos.

- Calculando directamente el valor de saliencia de cada segmento.

El segundo método es el más ventajoso, ya que podemos lograr una estimación de la saliencia más precisa al considerar las relaciones de segmentos u objetos en lugar de píxeles a nivel global. Esto es así, puesto que el ojo humano establece niveles de saliencia por conjuntos que guardan relación, por lo que es más beneficioso calcular la saliencia asociando píxeles.

Por esta razón se decanta por hacer una segmentación previa y optar por la segunda de las dos formas posibles.

En la sección 3.2 introduciremos la herramienta de segmentación utilizada por Muratov incluyendo un estudio y la adaptación de sus parámetros de configuración. La sección 3.3 tratará de explicar la técnica de saliencia utilizada por Muratov y analizará la calidad de la saliencia obtenida para diferentes imágenes propuestas.

## 3.2 Segmentación

Muratov hace uso de la herramienta de segmentación de EDISON desarrollado por *Robust Image Understanding Laboratory at Rutgers* basándose en [6], [23] y [24], siendo EDISON la abreviatura de *Edge Detection and Image SegmentatiON*.

Para poder utilizar la herramienta con Matlab deberemos compilarla con ‘*mex*’ mediante un compilador de ‘*C*’.

Esta herramienta tiene como parámetro principal de entrada la imagen de la que queremos obtener su mapa de saliencia, además de pasarle una función que transforma de formato RGB a formato Lab. Después se pueden añadir unos parámetros fijos establecidos que se pueden variar según nos interese, y que de hecho han sido modificados. Estos parámetros son los siguientes:

- ‘SpatialBandwidth’ especifica el tamaño de la ventana de búsqueda para el algoritmo de mean-shift. El tamaño de la ventana viene determinado por  $(2r+1)*(2r-1)$  donde *r* es el ‘SpatialBandWidth’.

- 'RangeBandWidth' especifica el tamaño de la ventana de búsqueda en el subespacio que nos encontramos durante el cálculo de la mean-shift.
- 'MinimumRegionArea' especifica el área mínima permitida (tamaño del segmento en píxeles) que figura en la imagen segmentada.
- 'SpeedUp' determina la velocidad de ejecución del algoritmo de mean-shift. Muratov lo deja predeterminado a 2.
- 'GradientWindowRadius' es el tamaño de la ventana utilizado para el cálculo del mapa de gradiente.
- 'MixtureParameter' especifica si le damos más peso al gradiente o a los bordes.
- 'EdgeStrengthThreshold' especifica el umbral para fusionar o no una zona con otra en función de sus bordes.

Minimum region area	$imheight \times imwidth \times 0.005$
Spatial bandwidth	10
Range bandwidth	7.5
Gradient window radius	2
Mixture parameter	0.3
Edge strength threshold	0.7

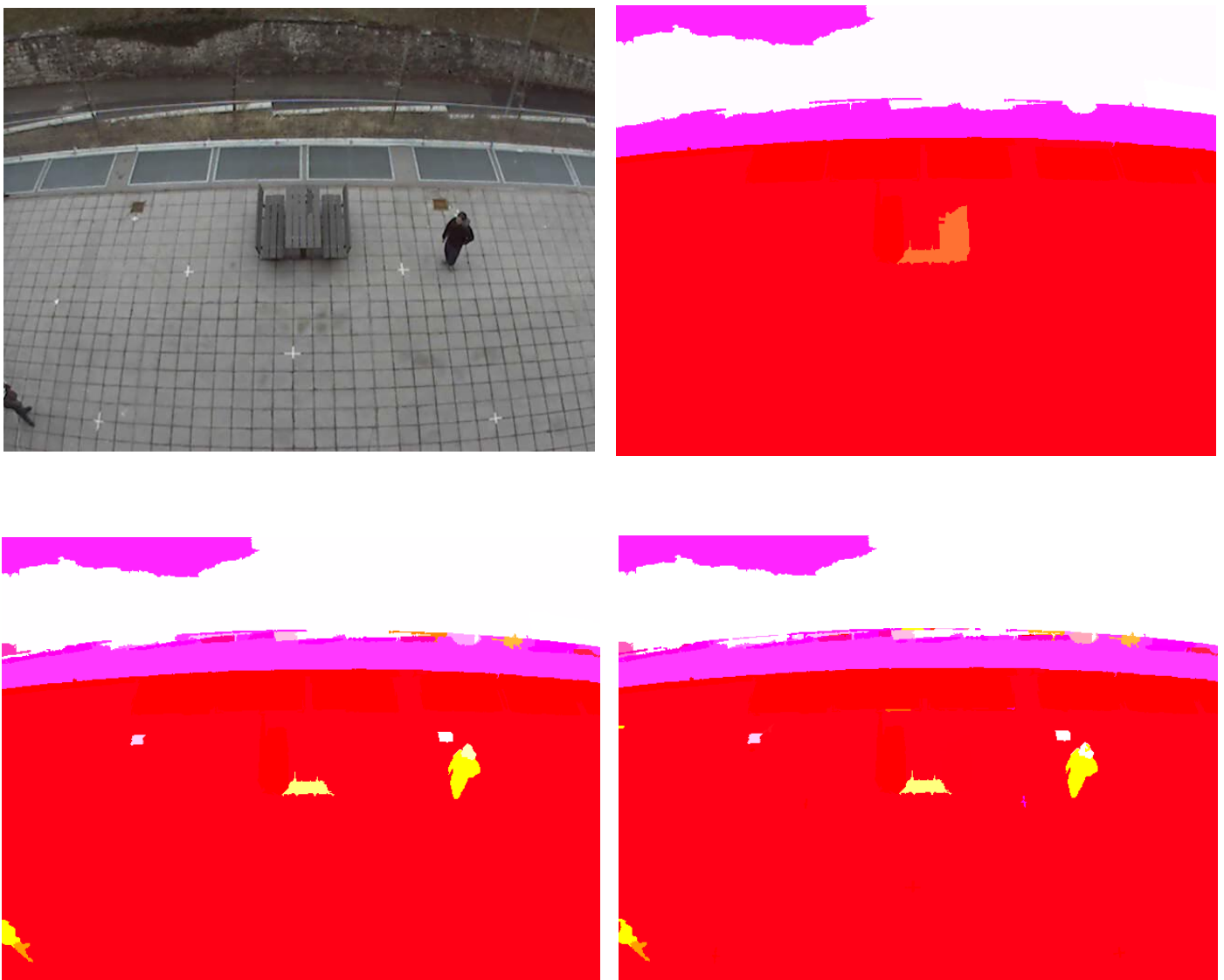
**Tabla 3 Parámetros EDISON por Muratov**

Después de diferentes pruebas se modificó el primer parámetro. El área de la región mínima que puede tener un segmento establecida por Muratov y que aparece en la Tabla 3, se calcula como un 0.5% del total de los píxeles de la imagen, lo que en planos de 640x480 corresponde a áreas de 1536 píxeles. El problema, es que cuando se establece la herramienta para un análisis de una secuencia de vídeo, el hecho de tener segmentaciones excesivamente grandes no nos permite detectar algunos objetos salientes. De igual modo, si elegimos áreas pequeñas con respecto al global de la imagen se produce inestabilidad temporal en la secuencia con segmentos que aparecen y desaparecen de un plano a otro por pequeñas alteraciones en los píxeles, en los que en realidad, y objetivamente, el ojo humano no realizaría distinción alguna. Por este motivo el parámetro modificado se estableció con un área mínima de

100 píxeles, dado que los vídeos con los que se ha trabajado tienen unas dimensiones aproximadas de 640x480 y recogen objetos interesantes con un tamaño aproximado a 100 píxeles. Es importante destacar que este parámetro deberá variarse en función de las características del tipo de secuencias de imágenes que se consideren y sería importante desarrollar un mecanismo para determinarlo de forma automática.

Minimum region area	100
---------------------	-----

**Tabla 4 Parámetro modificado**



**Figura 9 Segmentación variando el área mínima del segmento**

En la Figura 9, la imagen del cuadrante superior izquierdo es el plano original, a su derecha estamos segmentando con el área establecido por Muratov, mientras que en el cuadrante inferior izquierdo se ha utilizado un área mínimo de 100 píxeles y en el cuadrante inferior derecho con un área de 10 píxeles.

Como se puede observar, se cumple que para segmentos demasiado grandes no se segmentan objetos que pueden ser de especial importancia para el algoritmo de saliencia que se aplica posteriormente sobre el mapa de segmentación, como es el caso de las personas.

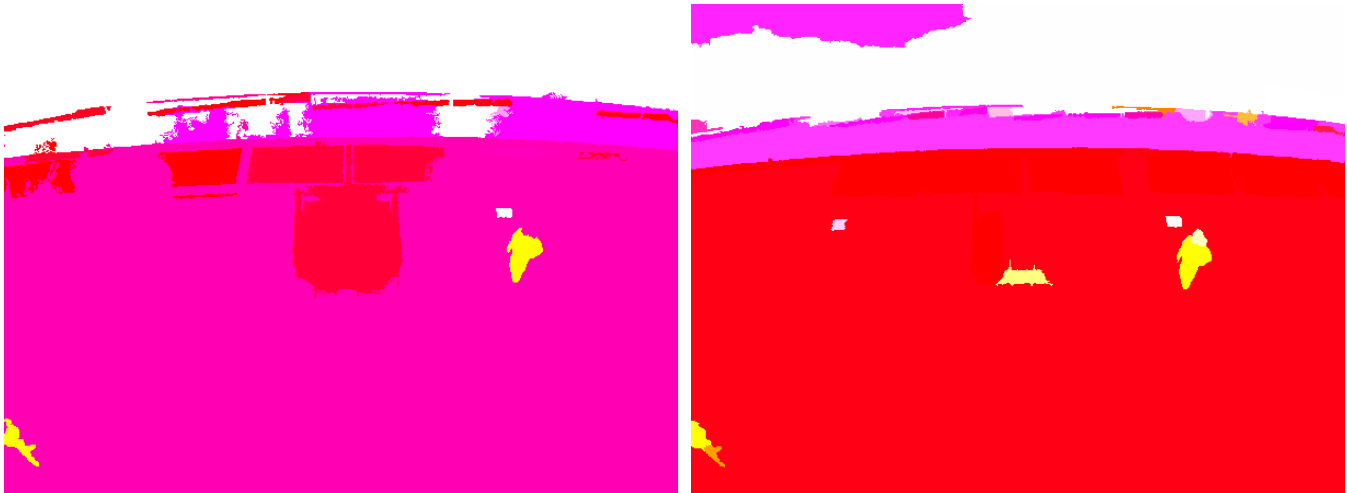
En el otro extremo, con un área muy pequeña se detectan segmentos que no tienen ninguna relevancia pero que afectan de igual modo al algoritmo de saliencia negativamente, por lo que es conveniente eliminarlos.

Ya que en la imagen puede que no se vean todos debido a su pequeño tamaño ofrecemos en la Tabla 5 los datos concretos del número de segmentos para cada caso.

Segmentación 1 Área 0,5% del global	10 segmentos
Segmentación 2 Área 100	43 segmentos
Segmentación 3 Área 10	113 segmentos

**Tabla 5 Número de segmentos para el mismo plano variando el área**

Otro parámetro que se modificó para intentar mejorar la salida segmentada por el problema de tamaño de los objetos mencionado anteriormente fue el 'SpatialBandwidth'. En este caso aumentándolo no se obtenían diferencias relevantes, pero el tiempo de ejecución del algoritmo aumentaba notoriamente. Se muestra en la Figura 10 y Tabla 6.



**Figura 10 Modificación del Spatial bandwidth**

Minimum region area	100	100
Spatial bandwidth	10	50
Elapsed time	5.862123 seconds	77.264905 seconds
Segmentos	43	31

**Tabla 6 Comparativa modificación Spatial bandwidth**

Por esta razón se desecha la opción de modificar el parámetro, el número de segmentos es semejante, pero el aumento del tiempo de ejecución sufre mucho.

Por el contrario si aumentamos el ‘Range bandwidth’ el número de segmentos disminuye bastante con poco que aumentemos su tamaño, por ejemplo para un *Range bandwidth* de 9,5 en lugar de 7,5 sin modificar el resto de parámetros, obtenemos 26 segmentos en lugar de 43 y aumentamos también el tiempo de ejecución en aproximadamente 1 segundo.

El ‘Gradient window radius’ ha de ser mayor que 0, permanece con valor 2 porque no tiene influencia notoria sobre el mapa de saliencia.

El 'MixtureParameter' y el 'EdgeStrengthThreshold' los mantenemos como Muratov puesto que luego trabaja con esos valores para algunos descriptores de la saliencia.

Mixture parameter	0.3
Edge strength threshold	0.7

**Tabla 7 Parámetros fijos Muratov**

### 3.3 Detección de saliencia

La herramienta de Muratov utiliza la salida obtenida de la segmentación para generar una serie de descriptores que posteriormente ponderarán en la salida del mapa de saliencia. También hace uso de la luminancia de la imagen original y un filtro basado en [21] denominado *Center-surround histogram* que genera un histograma situando una ventana de análisis encima de cada segmento y analiza la saliencia para ese segmento de forma que ese histograma y su saliencia sean únicos para ese segmento.

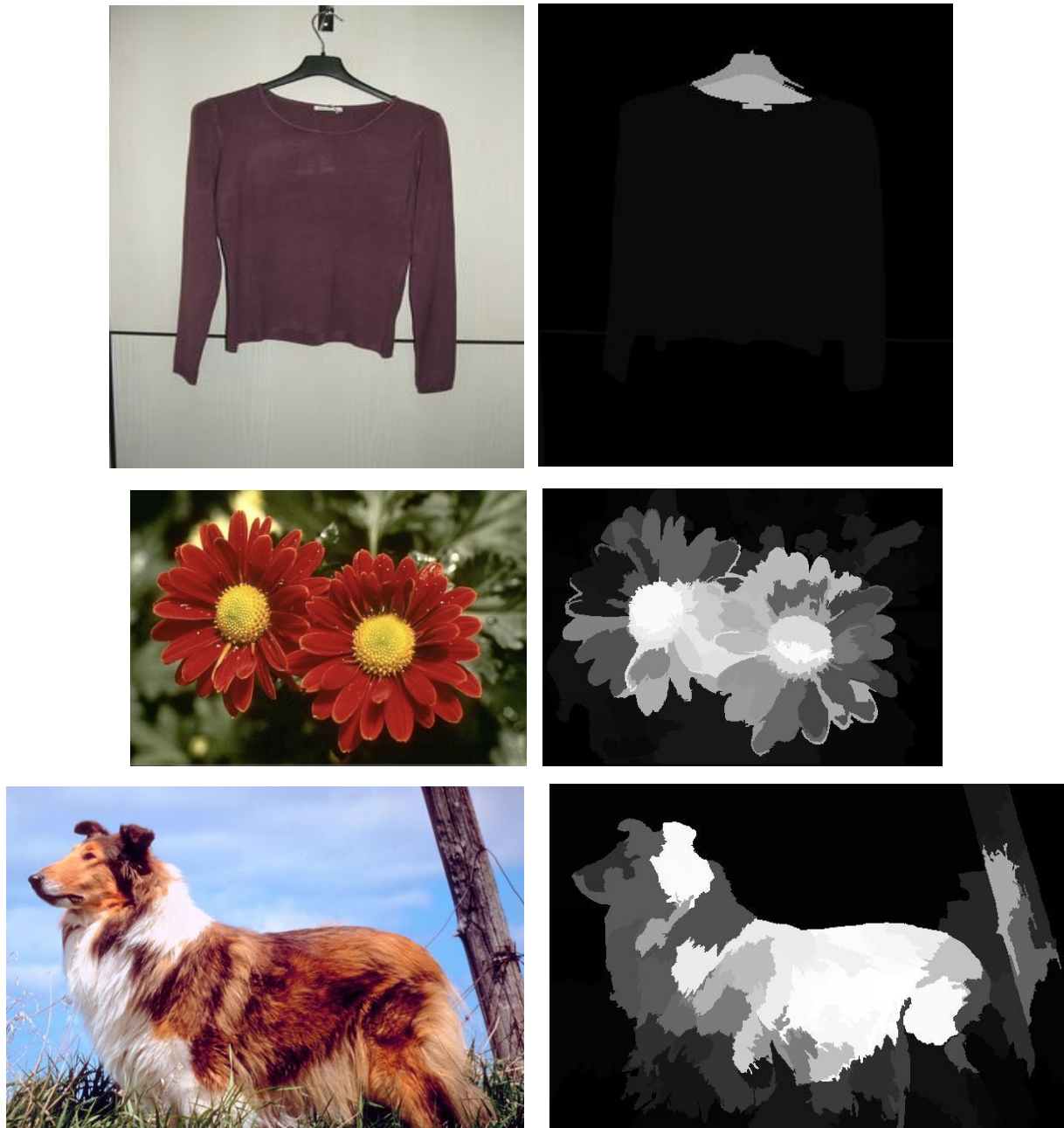
Los descriptores se basan en el tamaño del objeto, su forma, su localización en el cuadro de la imagen, la dominancia de los colores, la distribución de la misma, las diferencias entre primer plano y fondo, es decir, qué parte ocluye la visión de otra. Por otra parte hace una clasificación basada en la geometría del segmento ya que los objetos que tienen esquinas o con vértices muy definidos suelen corresponder a cielos o suelos.

Los valores ofrecidos por los descriptores pasan a un clasificador de tipo Naïve Bayes, que es un clasificador probabilístico basado en el teorema de Bayes junto a la hipótesis de que las variables predictoras son independientes utilizando el método de máxima verosimilitud y por lo que al ser independientes sólo es necesario estimar los parámetros de media y varianza de las variables y no toda la covarianza. Este tipo de clasificadores requiere menos volumen de datos de entrenamiento para estimar los parámetros de máxima verosimilitud.

Finalmente hace uso de la salida del clasificador bayesiano junto a la segmentación obtenida con EDISON para generar el mapa de salida de saliencia.

A continuación se muestra en la Figura 11 la saliencia de diferentes imágenes prestadas por la Universidad de Trento, para comprobar si se trata de un modelo cuya salida concuerde cualitativamente con la percepción de saliencia visual humana. Como se trata de algo subjetivo se ha pedido la colaboración de varias personas.

Las zonas más salientes se representan por grises claros y las menos salientes por grises oscuros.



**Figura 11 Ejemplos de Saliencia Muratov**



Sólo se ha evaluado la zona más saliente y para ello la pregunta realizada ha sido: ¿qué zona de la imagen te resulta más llamativa a primera vista?

Como se muestra en la Tabla 8, para el primer caso el algoritmo aprueba en un 50% de los casos, para el segundo caso en un 90% y para el tercer caso un 60% por lo que con este pequeño análisis se puede deducir que el algoritmo aprueba en aproximación a la realidad.

Sujeto	Imagen 1	Imagen 2	Imagen 3
1	SI	SI	NO
2	NO	SI	SI
3	SI	SI	SI
4	SI	SI	NO
5	NO	SI	SI
6	NO	SI	NO
7	SI	SI	SI
8	NO	NO	NO
9	NO	SI	SI
10	SI	SI	SI

**Tabla 8 Encuesta sobre saliencia**

# Capítulo 4

## Adaptación de la herramienta de Muratov para secuencias de vídeo

### 4.1 Introducción

La herramienta de detección de saliencia de Muratov está pensada para ejecutarla sobre una sola imagen. En este proyecto hemos buscado ampliar la saliencia de una imagen estática a una secuencia de imágenes y para ello hemos tenido que realizar varios cambios que afectan tanto a la ejecución de la herramienta cómo a su algoritmo.

Para que se pueda apreciar el trabajo llevado a cabo en esta tesis y puesto que las medidas objetivas no son posibles por su propia naturaleza, hemos dejado ejemplos de los resultados disponibles en YouTube.

El vídeo utilizado es una grabación de una cámara de seguridad procedente de una base de datos de la Universidad de Trento con el cual se llevaron a cabo experimentos de seguimiento visual o *eye-tracking* de los que pretendíamos hacer uso para entrenar el algoritmo. En la grabación se presenta una plaza donde hay viandantes que la cruzan de un lado a otro y también

otros que están parados. El vídeo completo puede verse en la siguiente dirección:

[ <http://youtu.be/hzLzvIKq9l0> ]

A lo largo de este capítulo mostraremos los procesos de adaptación de la herramienta de Muratov en la sección 4.2, así como los problemas con los que nos hemos encontrado en la sección 4.3 y las propuestas, soluciones y alternativas de diseño que se han utilizado para resolverlos en la sección 4.4. Por último hemos incluido una sección de resultados.

## 4.2 Adaptación

Para que se pudiese realizar un análisis de una secuencia de vídeo se ha realizado un pequeño código que separa los planos de un vídeo y los guarda como imágenes independientes para posteriormente pasárselas como parámetro a la herramienta de saliencia y que esta genere el mapa de saliencia almacenándolo en un directorio. Posteriormente se monta de nuevo el vídeo plano a plano con los mapas de saliencia.

Por resultarnos de interés observar la salida del segmentador EDISON existe la posibilidad de guardar la salida del segmentador e incluso poder realizar un procesado de esta salida y devolverlo al mismo punto para que continúe el proceso y obtenga el mapa de saliencia.

## 4.3 Problemas encontrados

Como ya lo referenciamos en las fases del Diagrama de Gantt de la Figura 1 en el primer capítulo sección 1.5, el principal problema que ha surgido en este punto se ha extendido en tiempo más de lo que pensábamos en un principio.

Se trata de un problema de estabilidad en la saliencia, esto quiere decir, que en la transición de un plano al siguiente en los mapas de saliencia y en los de la segmentación no se mostraban unas modificaciones razonables, es decir, no existía una transición suave en planos consecutivos cuando los datos de

entrada (las imágenes) que pasan por parámetro a la herramienta de Muratov guardan una correlación elevada. Se puede observar en la Tabla 9.

0,99636111	0,99687121	0,99692474	0,99711215	0,99714857	0,98766154	0,99202428
0,99716015	0,99710999	0,99684682	0,99695747	0,99652835	0,98706536	0,98595945
0,99691826	0,99437667	0,9955818	0,99662072	0,99592184	0,98558019	0,98490207
0,99675404	0,99611624	0,99611987	0,99669431	0,99664367	0,9807044	0,9884901
0,99572211	0,99563162	0,99666832	0,99427421	0,99452657	0,99097377	0,98765316
0,99603901	0,99502405	0,99635892	0,99633814	0,99299139	0,98944168	0,98823097
0,99529586	0,9954153	0,99352513	0,99608401	0,99248202	0,98945308	0,98542011
0,99176837	0,99469006	0,9946442	0,99463087	0,98972089	0,9901914	0,99275474
0,99507013	0,9946889	0,99050617	0,99421334	0,98931177	0,9916256	0,99063438
0,99354108	0,99398504	0,99188371	0,98938585	0,99317234	0,98776454	0,9931417
0,98502184	0,99312404	0,99294221	0,99262802	0,98810304	0,98690869	0,98941463
0,99306111	0,99232232	0,99208759	0,98789374	0,98985836	0,98861703	0,99216702
0,98326093	0,98773228	0,99324183	0,98933432	0,99373088	0,99008384	0,98794532
0,99391084	0,99002613	0,99373994	0,99350668	0,98920212	0,988355	0,98632723
0,98941211	0,98660346	0,98815181	0,99295559	0,99320567	0,99255878	0,99381237
0,99327201	0,98869528	0,99278964	0,98790909	0,99271298	0,98935733	0,99147191
0,98462209	0,98855123	0,99362174	0,99053125	0,9912947	0,99213822	0,99211666
0,99298626	0,9881664	0,98729087	0,99206492	0,99163232	0,99255919	0,99221946
0,98910038	0,99092519	0,99203522	0,98816751	0,99262921	0,99184626	0,99230114
0,98511382	0,98545603	0,99132975	0,9910239	0,98531239	0,98994528	0,98519987
0,99366864	0,98547961	0,99177109	0,98504905	0,98494551	0,99233192	0,98681844
0,99167589	0,9895738	0,98756529	0,98229908	0,98950313	0,98848042	0,99340575
0,98835516	0,98282668	0,98991516	0,98993739	0,97873507	0,99196196	0,99239055
0,98723462	0,98490331	0,98679814	0,99209336	0,9882765	0,99001206	0,99250864
0,99018872	0,9849301	0,98470676	0,98662321	0,99206727	0,99034139	0,98841022
0,98797698	0,99158142	0,98748887	0,98397285	0,99034496	0,98719382	0,98893889
0,98809392	0,99054102	0,99098073	0,98652255	0,98793259	0,99145684	0,98904795
0,99248793	0,99262332	0,98562073	0,99237061	0,98905536	0,98865963	0,99194744
0,98944674	0,99216838	0,98597769	0,99197146	0,99220884	0,98551074	0,98644085
0,9936417	0,9897381	0,98890849	0,99276021	0,99255538	0,98809348	0,98703007
0,99105082	0,98517013	0,99093038	0,99192797	0,98693963	0,99259249	0,98791076
0,99268171	0,99263068	0,99293604	0,98853187	0,98862963	0,99324988	0,99028601

**Tabla 9 Correlación entre planos consecutivos**

Esto ha motivado una serie de propuestas y soluciones con sus correspondientes alternativas de diseño, ya que a la larga, ha pasado de ser una fase a ser el objetivo principal para este proyecto, en el que se han investigado las posibles y diferentes causas que podrían estar afectando a la correlación temporal en la salida de saliencia obtenida.

Hemos tenido otra serie de problemas que han surgido al buscar una solución al problema principal del que estamos hablando y que se explicarán en la sección 4.4 según han ido apareciendo, puesto que es más sencillo encontrarles sentido desde el momento en que afloran.

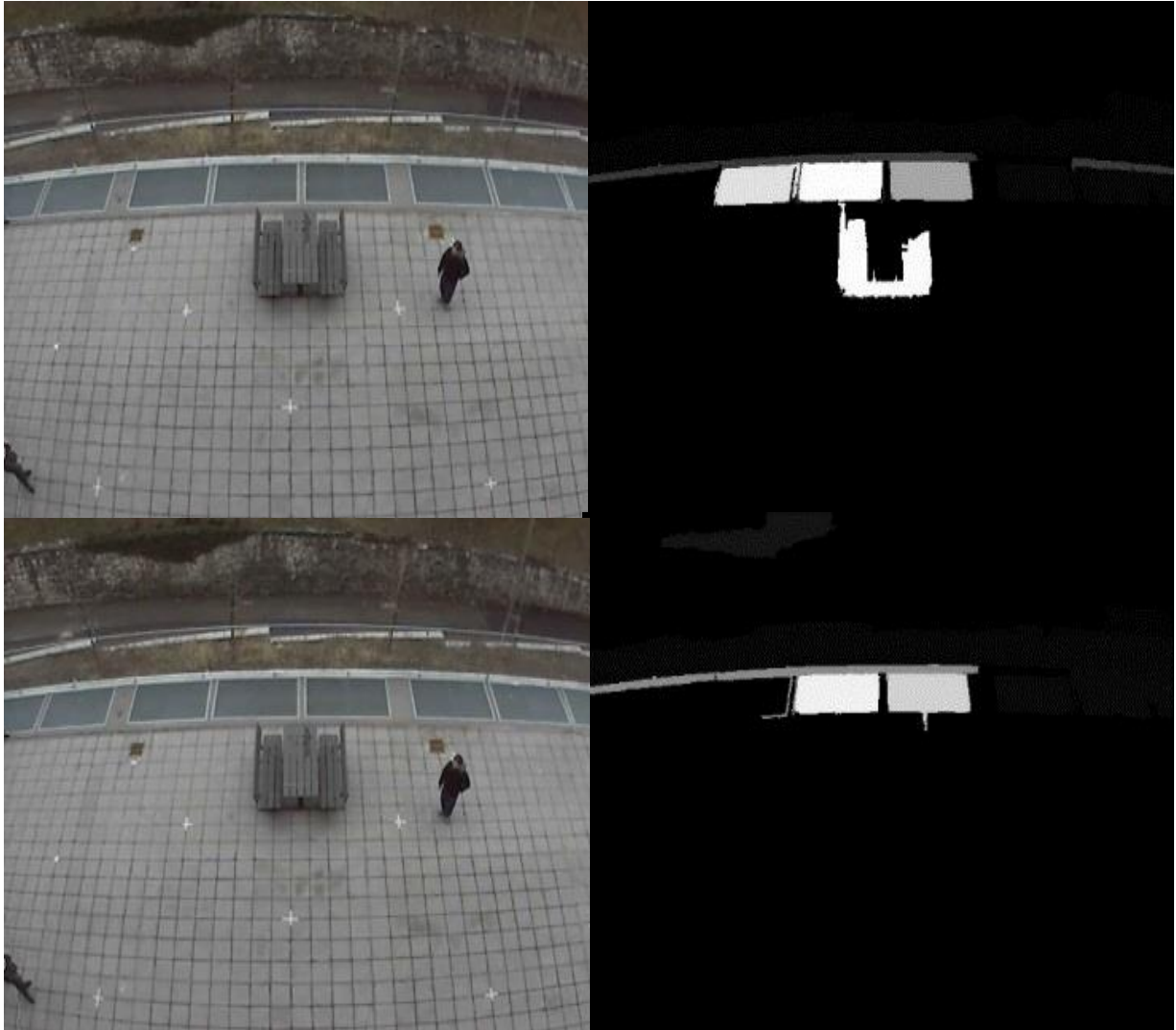
## **4.4 Propuestas, soluciones y alternativas de diseño**

Primero explicaremos las propuestas que llevamos a cabo pero que no dieron solución al problema expuesto en la sección anterior y de otras soluciones a problemas que fueron surgiendo para, posteriormente, hablar de la solución que dimos al problema principal y las posibles alternativas de diseño que podemos aplicar para la solución descrita.

En un principio pensamos que podría tratarse de una inicialización aleatoria en algún punto del algoritmo de saliencia, porque no era razonable que de un plano al siguiente, cuando no se habían producido cambios significativos en los píxeles, se pudiese generar una salida tan diferente de uno a otro. Pero tras llevar a cabo una serie de pruebas con el mismo plano se pudo comprobar que la salida para este era siempre la misma, por lo que quedó desestimado.

Posteriormente creímos que podría tratarse de un problema en el bajo nivel, en la segmentación, puesto que como explicamos en el Capítulo 3, sección 3.1, Muratov trabaja con imágenes segmentadas previamente con otra herramienta de segmentación denominada EDISON. Esta herramienta por alguno de los parámetros seleccionables u otras razones podría estar llevando a cabo segmentaciones diferentes en cada caso.

Lo primero fue analizar los segmentos de salida de cada plano y llegados a este punto, se comprobó que de un plano a su consecutivo, la segmentación obtenida sufría grandes cambios, existía a simple vista como se puede observar en la Figura 12 poca correlación entre planos consecutivos y este podría ser el problema que afectase en gran medida a la estabilidad en la salida de la saliencia.

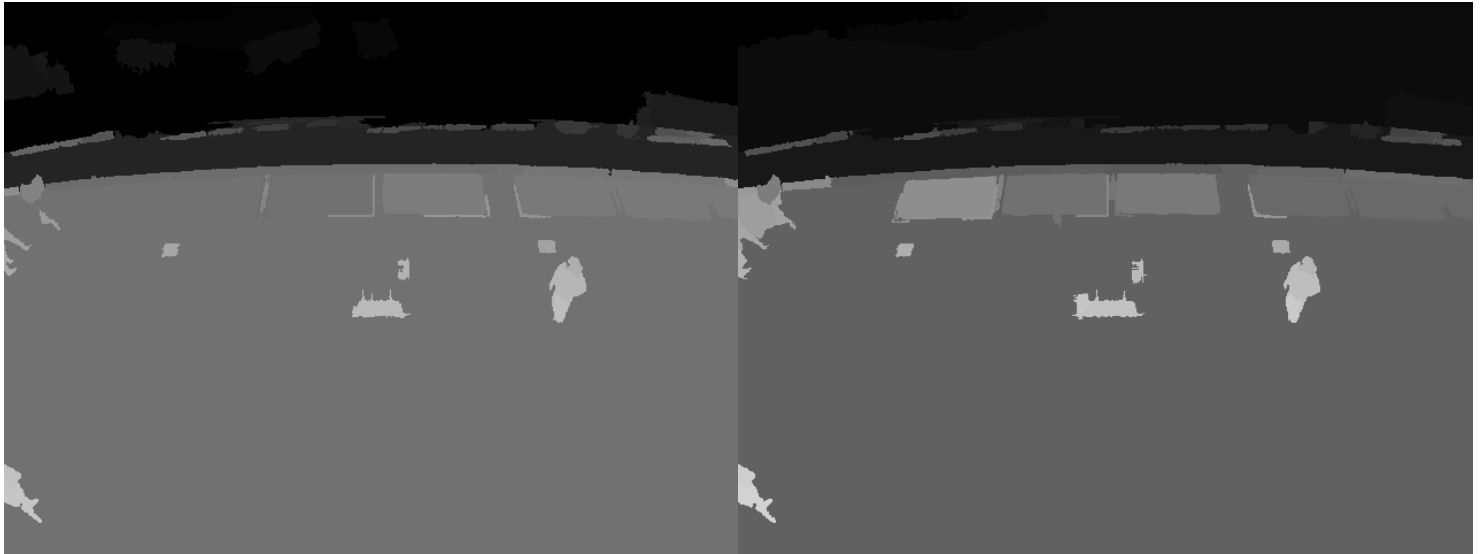


**Figura 12 Imagen Vs Imagen Segmentada en planos consecutivos**

Este problema ha dado lugar a escribir una sección sobre ello, en la sección 3.2, en la cual se aporta el estudio y las soluciones y en este capítulo solo se menciona puesto que tiene más relación con el análisis de la segmentación en la herramienta de Muratov, aunque es necesario aludir a ello en este capítulo puesto que en un momento dado supusimos que limitándonos a corregirlo solventaríamos el problema de la inestabilidad, cosa que como ya se está dando a entender no fue suficiente.

La herramienta de EDISON ofrece una salida de *labels* o etiquetas numéricas que se corresponden con los segmentos de la imagen. Esto también generaba inestabilidades a la hora de representar dichas imágenes segmentadas (ya que han sido representados en escala de grises de menor a mayor según la numeración, esto es, el 1 con el negro y el 256 con el blanco), puesto que la asignación de etiquetas es arbitraria, siempre en orden creciente

de izquierda a derecha del plano y de arriba abajo del mismo. Esto hace que si en un plano tenemos un segmento situado más a la izquierda de otro y/o más arriba del plano y en el siguiente plano éste desaparece por cualquier motivo, el etiquetado cambia produciendo salidas diferentes a la hora de representarlo como se puede observar en la Figura 13, en la que el plano de la izquierda tiene uno de los fondos del suelo en un tono de gris más claro que el de la derecha.

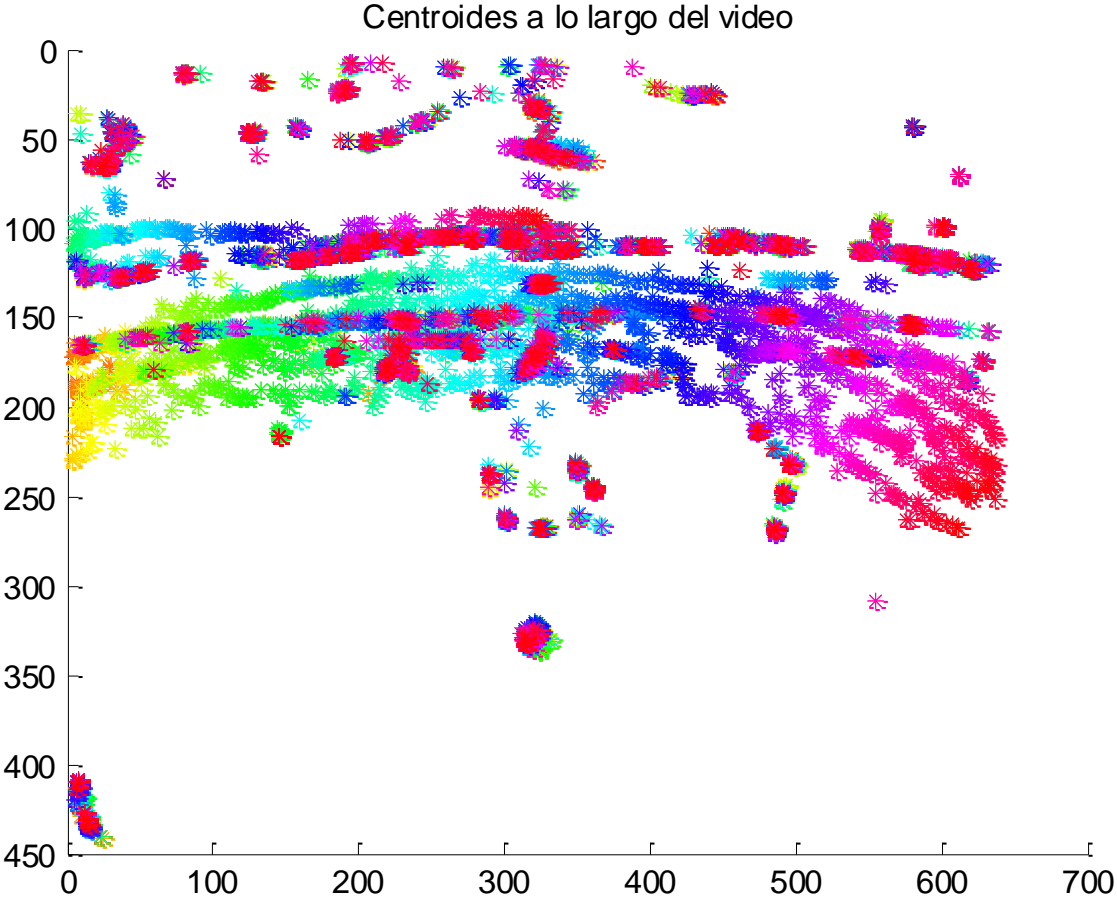


**Figura 13 Variación en la saliencia para el mismo segmento**

Por esta razón se pensó analizar este problema y como solución se propuso un reordenamiento para poder realizar un seguimiento temporal de los segmentos a lo largo de la secuencia de vídeo. Primero se calcularon los centroides de cada uno de los segmentos y el área de cada uno de estos. En segundo lugar se llevó a cabo un algoritmo de ordenación de los segmentos en función a estas características tomando como valores fijos los obtenidos del plano intermedio de la secuencia de vídeo para nuestro caso.

El vídeo sobre el que se ha trabajado, es el que hemos venido utilizando hasta ahora, el de la cámara de seguridad, en el que se van desplazando una serie de personas.

A continuación, en la Figura 14 se muestran los diferentes centroides en cada plano de dicho vídeo en una misma gráfica. El color indica el plano, los naranjas-amarillos son los primeros y los violáceos-rojos los últimos planos del vídeo.

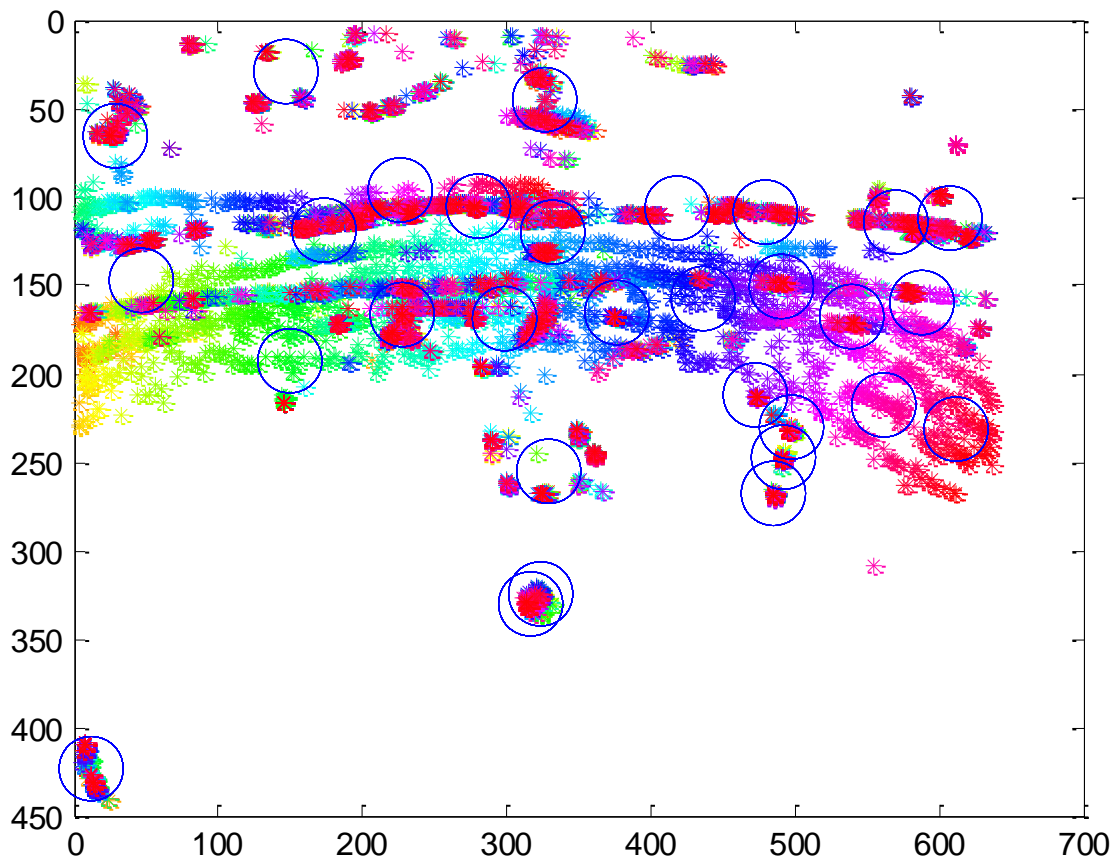


**Figura 14 Cálculo de los centroides**



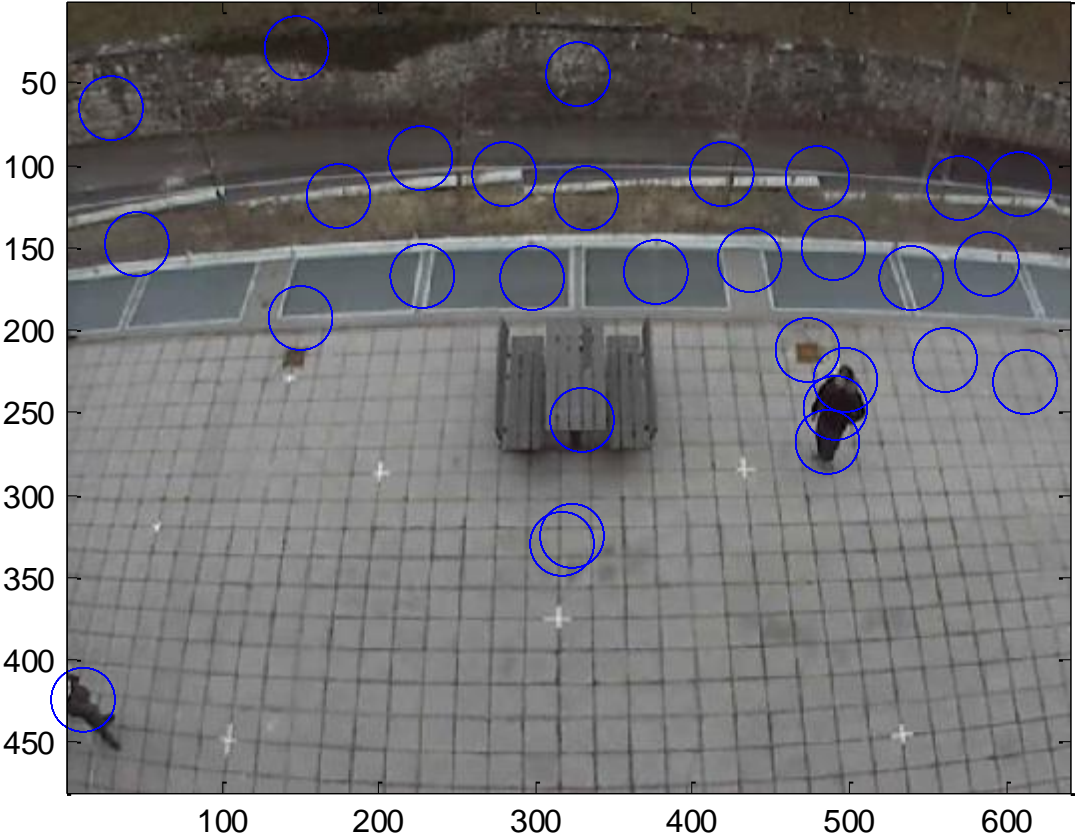
Como se puede observar en la Figura 15, se ha añadido el cálculo de los 20 k-means más relevantes, que se representan remarcados con un círculo azul. Esto nos indicaría los segmentos que permanecen más tiempo en la misma posición, mientras que el resto son el resultado de objetos que se mueven a lo largo de la secuencia de vídeo.

Se puede observar el flujo de objetos existente en cierta medida representado por aquellos puntos que quedan fuera de los círculos azules.



**Figura 15 K-means sobre centroides**

En la Figura 16 se puede comprobar que los segmentos-objetos fijos del vídeo quedan aceptablemente marcados por el k-means. Si vemos el vídeo utilizado podremos comprobar que los círculos azules se corresponden en gran medida con objetos estáticos en el plano. Para tener un mejor resultado del algoritmo que calcula el k-mean conviene realizar la media de varias ejecuciones, puesto que trabajan con heurísticas que pueden dar lugar a diferentes soluciones.



**Figura 16 Comparativa K-means con plano original**

Pese a este buen resultado no obtuvimos una mejoría con el algoritmo de reordenamiento para lograr obtener una mayor estabilidad. Habíamos aplicado demasiado tiempo en este punto, por lo que lo desestimamos como posible solución y dejamos de lado la propuesta de llevar a cabo un seguimiento temporal de los objetos para trabajar desde otro punto de vista. No obstante se podría mejorar en un futuro e incluso utilizarlo como un bloque paralelo al cálculo de la saliencia para obtener mejores resultados en un contexto global de la detección de la saliencia.

Por último se propuso un filtrado morfológico para eliminar el posible ruido existente en los planos, tanto espacial, dentro de cada plano, como temporal, para asignar una correlación temporal a los mismos. Esta fue la solución definitiva sobre la que se explicarán a continuación las posibles alternativas de diseño.

Llegados a este punto tenemos que hablar de filtrado morfológico tal y como explica González. R *et al* [25]. Existen dos operaciones morfológicas principales que pueden conjuntarse en diferente orden y que en nuestro caso representan las posibles alternativas de diseño de los filtros. Puesto que unas u otras maneras de llevarlo a cabo pueden ser más o menos interesantes según el problema que se nos presente, en nuestra situación hemos elegido la que más nos ha convenido, pero recalcamos que existe la posibilidad de la alternativa en el diseño.

A continuación se explican las dos operaciones y sus posibles relaciones:

- **Erosión**

Operación en la cual dada una imagen y un elemento estructural, la erosión de la imagen por el elemento estructural viene dada por el conjunto de todos los píxeles de la imagen para los cuales el elemento estructural trasladado por los píxeles está contenido en la imagen.

$$A \ominus B = \{x \mid B_x \subseteq A\}$$

Cuando aplicamos esta operación obtendremos una reducción del tamaño de los segmentos de la imagen que se vean afectados por el elemento estructural, por lo que el tamaño del elemento estructural cobra una importante relevancia en la operación. A continuación, en la Figura 17 se muestra un ejemplo de la operación.

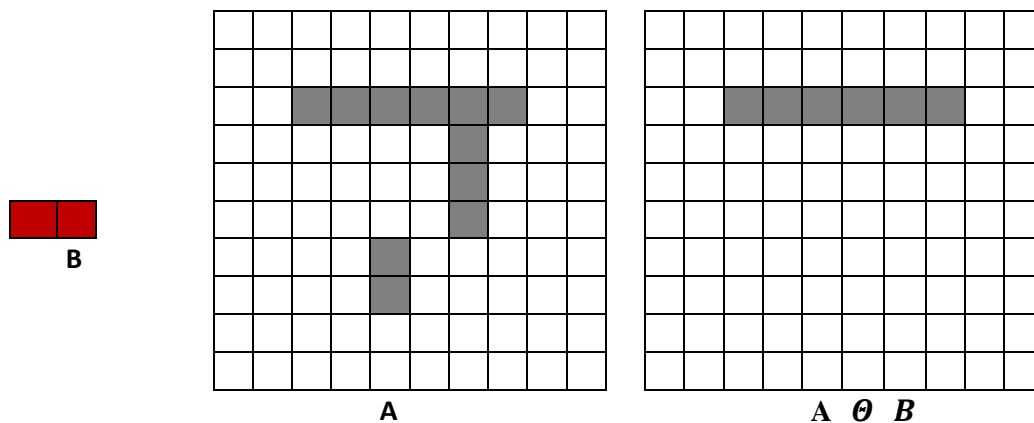


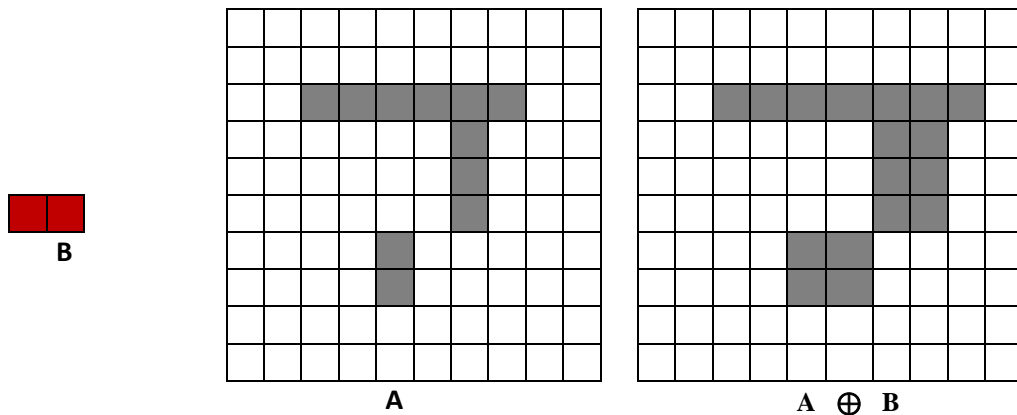
Figura 17 Erosión

- **Dilatación**

Operación en la cual dada una imagen y un elemento estructural, la dilatación de la imagen por el elemento estructural viene dada por el conjunto de puntos del elemento estructural tal que al menos contenga algún pixel cuando el elemento estructural es trasladado por los píxeles de la imagen.

$$A \oplus B = \{x | (\hat{B})_x \cap A \neq \emptyset\}$$

Cuando aplicamos esta operación obtendremos una ampliación de los segmentos de la imagen que se vean afectados por el elemento estructural pudiendo unirse varios segmentos en uno único, por lo que el tamaño del elemento estructural cobra una importante relevancia en la operación. A continuación, en la Figura 18 se muestra un ejemplo de la operación.



**Figura 18 Dilatación**

- **Composición**

Si operamos conjuntamente la dilatación con la erosión, obtendremos diferentes salidas en función de qué operación realicemos en primer lugar, puesto que no cumplen la propiedad conmutativa. Si realizamos primero un proceso de erosión y posteriormente una dilatación del resultado obtenido, lo denominaremos en conjunto proceso de apertura, mientras si por el contrario, primero realizamos una dilatación y posteriormente una erosión denominaremos al proceso como cierre. Un punto importante a tener en cuenta es que ambas operaciones deben realizarse con el mismo elemento estructural.

En nuestro caso hemos llevado a cabo un proceso de cierre, puesto que nos interesaba disminuir el número de segmentos de la imagen y de esta manera aquellos segmentos cercanos que tienen un tamaño pequeño se unen a segmentos de mayor tamaño que los absorben.

$$A \bullet B = (A \oplus B) \ominus B$$

La combinación que da lugar al cierre es mejor que sólo dilatar la imagen que distorsionará todos los segmentos indiscriminadamente,

puesto que con el cierre se consigue minimizar este efecto ya que posteriormente erosionamos esa distorsión.

La posible alternativa sería realizar un proceso de apertura con la que obtendríamos más segmentos, en caso de un vídeo con pocos segmentos.

$$A \circ B = (A \ominus B) \oplus B$$

Como ya se ha mencionado, se propuso un filtrado morfológico para eliminar el posible ruido existente en los planos, tanto espacial, dentro de cada plano, como temporal para asignar una correlación temporal a los mismos.

A continuación se explica cómo se ha procedido en cada uno de los casos:

- **Filtrado espacial**

Para el filtrado espacial se ha utilizado un elemento estructurante circular de 20 píxeles que barre cada plano de 640x480 píxeles; primero realizando una operación de dilatación y posteriormente una de erosión. Se ha aplicado a la salida del segmentador EDISON sobre el mapa de etiquetas de los segmentos.

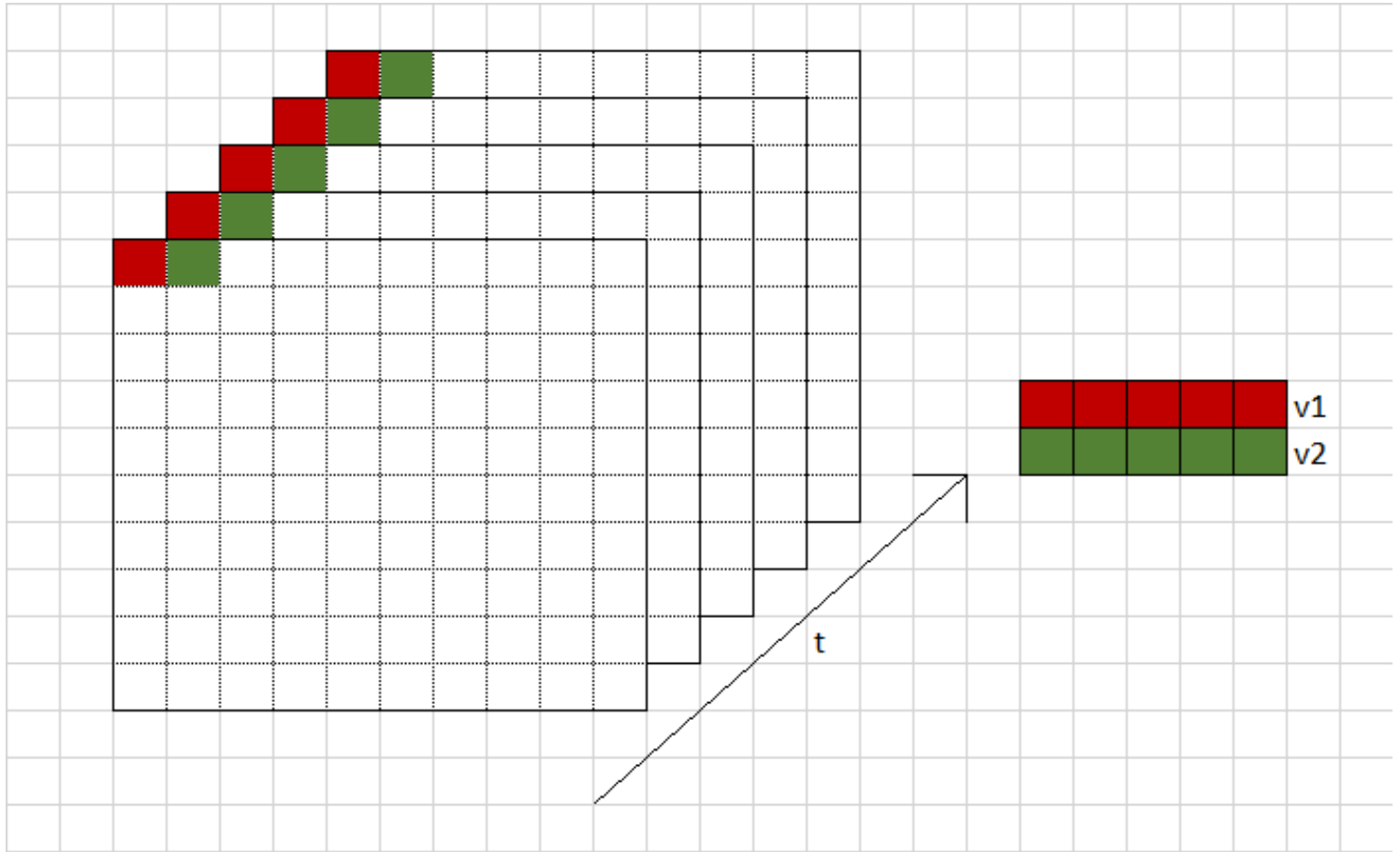
- **Filtrado temporal**

Para el filtrado temporal en primer lugar se cargan todos los mapas de etiquetas que fueron previamente filtrados espacialmente y generamos los vectores espaciales pixel a pixel a partir de ellos.

Llegado este punto tenemos 640x480xn vectores, donde n representa el número de planos de la secuencia de vídeo y es en cada uno de ellos en los que se realizan las operaciones de dilatación y erosión.

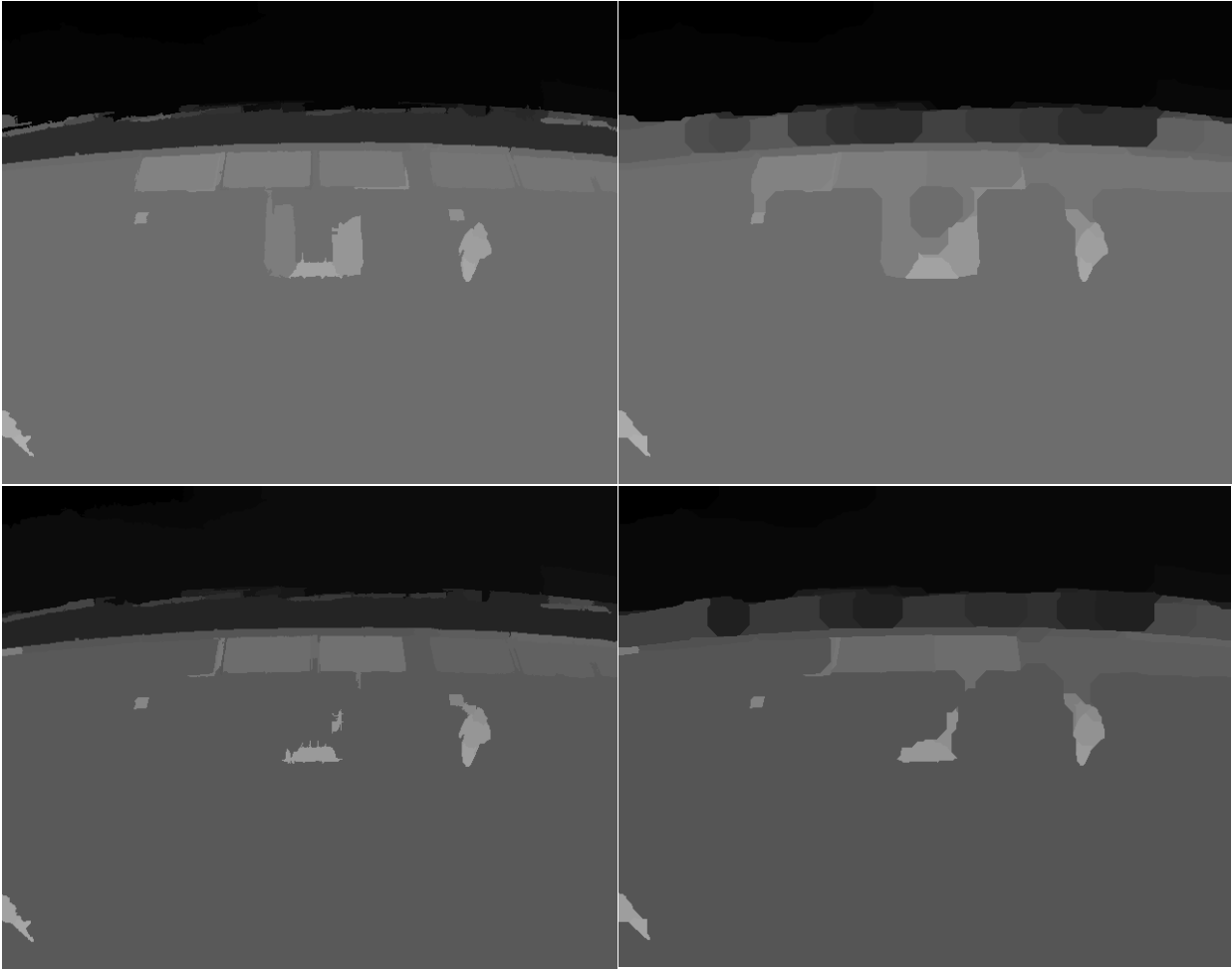
En el caso del filtrado temporal, utilizaremos el único elemento estructurante posible, que es un vector, una máscara de 1's de longitud máxima el tamaño del contexto elegido, en nuestro caso en todos los planos, pero si el vídeo es de extensa duración, con un número elevado de

planos, puede disminuirse el uso de la memoria eligiendo un filtrado espacial en un contexto más pequeño de planos. Se muestra en la Figura 19.



**Figura 19 Filtrado temporal**

En la Figura 20 se muestra una comparación entre la segmentación obtenida para dos planos consecutivos sin aplicar ningún filtrado (en la parte de la izquierda de la imagen) y la segmentación obtenida aplicando posteriormente el filtrado espacial de cierre con un elemento estructural circular de 20 píxeles.



**Figura 20 Comparación aplicando cierre sobre la segmentación de dos frames consecutivos**

Se puede comprobar cómo en ambos casos, se produce una absorción del segmento más pequeño por el más parecido cercano al de mayor tamaño.

El vídeo completo de la segmentación sin aplicar ningún filtrado se puede encontrar en el siguiente link:

[\[http://youtu.be/q8Uj-iz\\_eMM\]](http://youtu.be/q8Uj-iz_eMM)

Y el vídeo completo de la segmentación aplicando el filtrado espacial en la siguiente dirección:

[\[http://youtu.be/ZjglP144U6M\]](http://youtu.be/ZjglP144U6M)

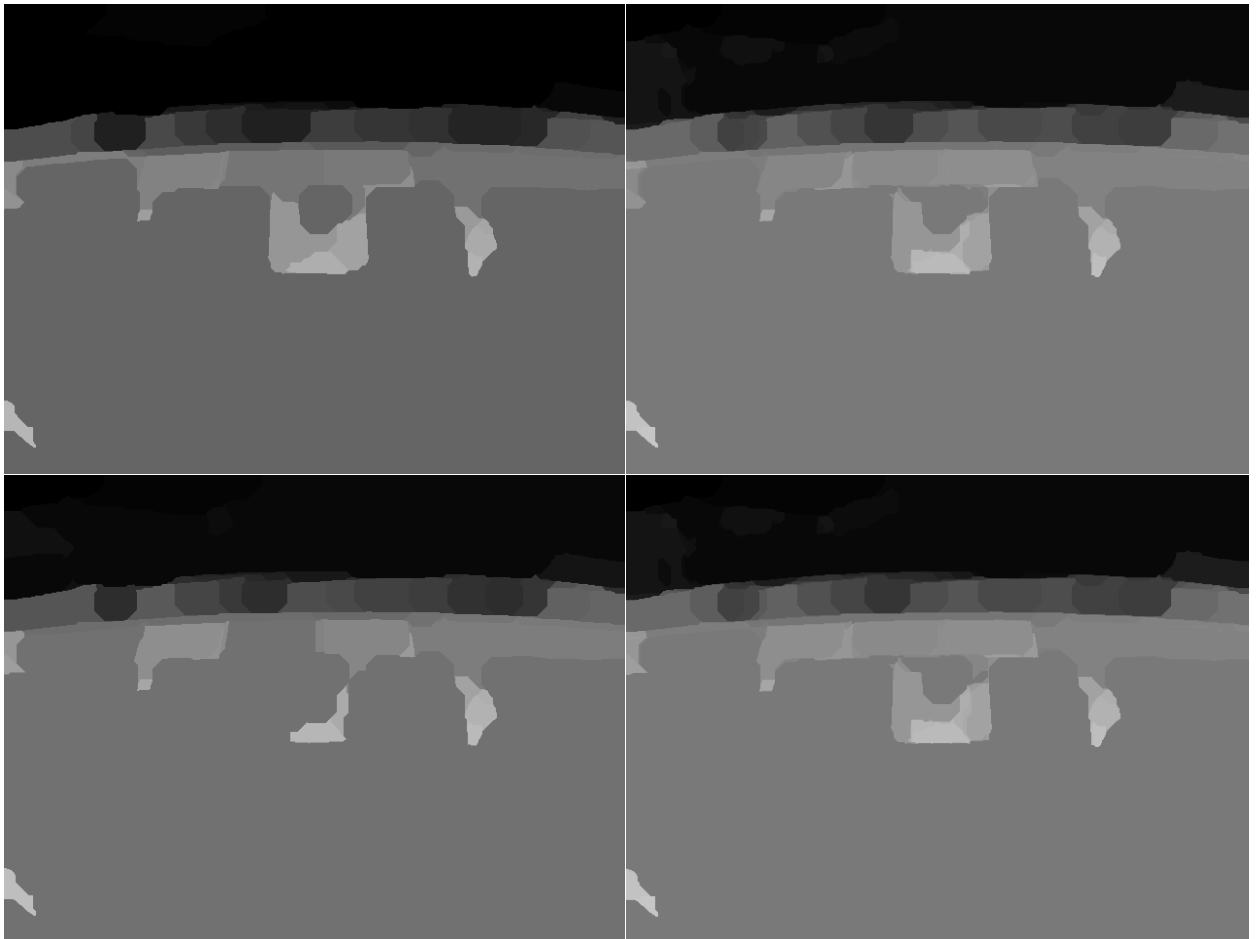
Además se puede observar que en ambos casos sigue sucediendo el problema de etiquetado, en el segmento que representa el suelo toma un gris



más oscuro o más claro según el plano, que era el problema que intentamos resolver con el seguimiento temporal de objetos, pero a continuación cuando apliquemos el filtrado temporal sobre el filtrado espacial el problema desaparecerá.

A continuación, en la Figura 21, se muestra una comparación para dos planos consecutivos entre la segmentación obtenida aplicando el filtro de cierre espacial (a la izquierda de la imagen) y la obtenida mediante un filtrado temporal sobre la salida del espacial, también en este caso aplicando un filtrado de tipo cierre. El filtrado espacial se realiza con un elemento estructurante circular de 20 píxeles mientras que el filtrado espacial se lleva a cabo con el siguiente elemento estructural vectorial:

$$[1 \ 1 \ 1 \ 1 \ 1]$$



**Figura 21 Comparación de filtrado en la segmentación**

El vídeo completo de la segmentación aplicando ambos filtros se puede encontrar en el siguiente link:

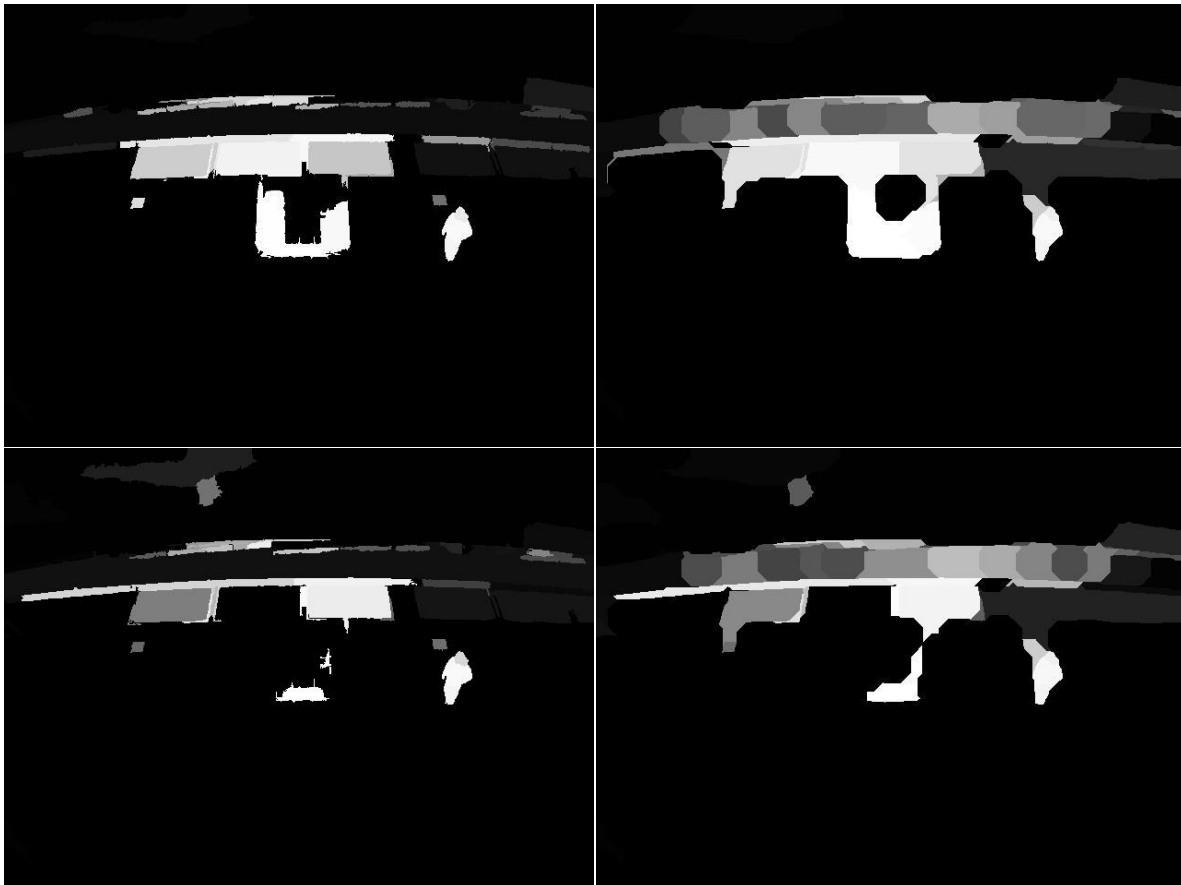
[<http://youtu.be/JZgu0QQ3itg>]

Ahora vamos a mostrar las salidas para la saliencia de los casos expuestos anteriormente para la segmentación.

El vídeo completo de la saliencia sin aplicar ningún filtrado se puede encontrar en el siguiente link:

[<http://youtu.be/rMN9pDcEKuU>]

A continuación se muestra en la Figura 22 una comparación entre la saliencia obtenida para dos planos consecutivos sin aplicar ningún filtrado (en la parte de la izquierda de la imagen) y la saliencia obtenida aplicando posteriormente el filtrado espacial.

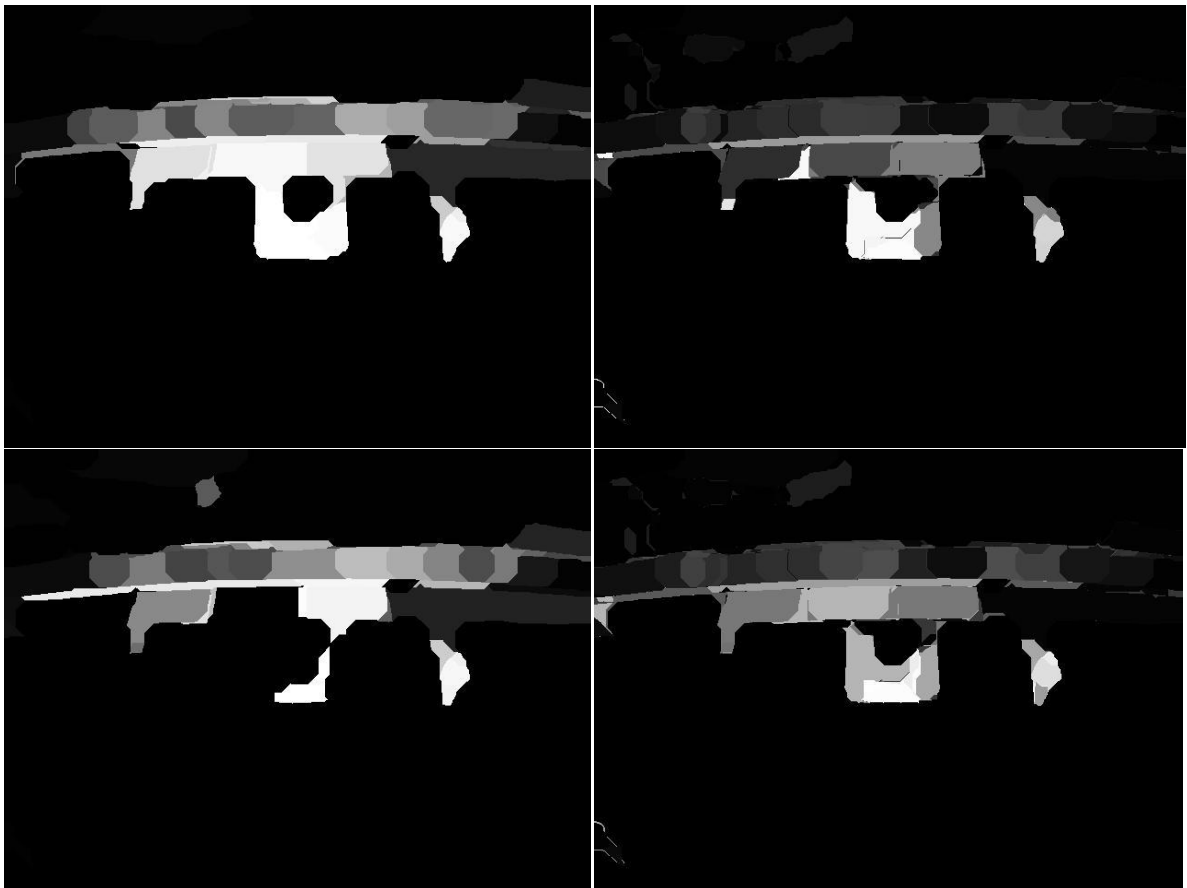


**Figura 22 Comparación aplicando cierre sobre la segmentación de dos frames consecutivos**

Y el vídeo completo de la saliencia aplicando el filtrado espacial en la siguiente dirección:

[\[http://youtu.be/KrsL7m6vuBo\]](http://youtu.be/KrsL7m6vuBo)

A continuación se muestra en la Figura 23 una comparación para dos planos consecutivos entre la saliencia obtenida aplicando el filtro de cierre espacial (a la izquierda de la imagen) y la obtenida mediante el filtrado temporal sobre el filtrado espacial.



**Figura 23 Comparación filtrado en Saliencia**

El vídeo completo de la saliencia aplicando los dos filtros se puede encontrar en el siguiente link:

[\[http://youtu.be/EOFQu1pmrhA\]](http://youtu.be/EOFQu1pmrhA)

## 4.5 Resultados

Para obtener un análisis numérico de los resultados y obtener así conclusiones objetivas, se ha decidido utilizar el estadístico de la correlación entre planos consecutivos y calcular una media del estadístico obtenido para cada par de planos. Los resultados se resumen en la Tabla 10:

	Correlación media de la secuencia de vídeo	Correlación mínima	Correlación máxima
Segmentación EDISON	<b>0.9849</b>	0.9537	0.9963
Segmentación con filtrado espacial	<b>0.9846</b>	0.9603	0.9959
Segmentación con filtrado espacial y temporal	<b>0.9987</b>	0.9857	1
Saliencia EDISON	<b>0.7841</b>	0.4196	0.9455
Saliencia con filtrado espacial	<b>0.7941</b>	0.5845	0.9446
Saliencia con filtrado espacial y temporal	<b>0.9639</b>	0.7865	0.9994

**Tabla 10 Resultados**

Como podemos observar, la correlación en la saliencia aplicando ambos filtrados aumenta considerablemente, lo que indica que hemos obtenido una salida más estable que cuando sólo aplicábamos la herramienta de Muratov, concretamente tenemos un 17,98% más de correlación.

El vídeo completo con las seis comparaciones se encuentra en el siguiente link:

[<http://youtu.be/CVz38iSrnq8>]

En la zona superior del vídeo se representan las salidas del segmentador y abajo las de la saliencia.

Hay que tener en cuenta que este estadístico nos sirve como indicador. En este caso porque el vídeo utilizado tiene una transición suave plano a plano, en otros casos donde existan transiciones bruscas tendríamos que buscar otro tipo de parámetros estadísticos.

# Capítulo 5

## Viabilidad

### 5.1 Entorno socio-económico

La saliencia de un elemento, ya sea un objeto, una persona o un píxel es el estado o cualidad por la que destaca con respecto a sus vecinos. La detección de la saliencia es un mecanismo clave para seleccionar el subconjunto más relevante de la información sensorial disponible.

Actualmente vivimos en un mundo completamente informatizado en el que se suben diariamente más de 150 horas de vídeo en diferentes servidores de Internet y es una de las formas más utilizadas a la hora de transmitir información, bien sea en la televisión o en otras plataformas web. Por ello se hace cada vez más necesario lograr satisfacer un modelo computacional que se asemeje en el mayor grado posible a la capacidad que tiene el sistema visual humano para discernir qué es lo relevante dentro de toda la información que se le transmite en un vídeo, que sepa qué información es relevante y cuál es redundante para desecharla o subestimarla.

Resulta muy interesante un modelo que lo satisfaga, puesto que los campos de aplicación son tan diversos como las posibilidades de mercado

de una herramienta de este calibre. Hoy en día, en el mundo, se celebran cientos de miles de eventos publicitarios de diferentes marcas, por no hablar de la enorme cantidad de anuncios en vídeo que se editan diariamente. ¿Cuántos de los publicistas y productores estarían dispuestos a pagar por una herramienta que les dijese antes de emitir su anuncio, si éste tiene o no capacidad para persuadir la atención visual del sujeto que lo está viendo con un número, con un dato muy próximo a la realidad? Un estudio lo llevaron a cabo H. Liu *et al.* [21], que utilizaron el mapa de saliencia para insertar publicidad en una imagen de la forma menos intrusiva posible y evitando obstruir el primer plano. Se muestra en la Figura 24



**Figura 24 Dónde resulta menos intrusiva la publicidad H. Liu et al. [21]**

Podemos viajar desde el mercado de la publicidad al mercado de sistemas inteligentes de seguridad o de la armamentística, hay un amplio registro.

Existen varios enfoques para la detección de saliencia estática, esto es, en imágenes. Aunque logran una salida de saliencia bastante aceptable aún hay margen para mejorarlo, tanto es así que se lleva en este proceso más de 20 años. Sin embargo hay un número mucho más reducido de investigaciones sobre saliencia en secuencias de imágenes/vídeo, lo que hace realmente interesante para el mercado una aplicación de este calibre.

## 5.2 Marco regulador

Actualmente no existe ningún marco regulador técnico, estamos ante los primeros pasos que se están dando en este ámbito y los resultados que ofrecen las aplicaciones de detección de saliencia no están suficientemente consolidados.

El primer paso a seguir debería ser una base de datos de libre acceso sobre la consecución o no de los objetivos de detección de saliencia para las herramientas más consistentes y tener de este modo una referencia menos subjetiva para poder realizar comparaciones con nuevas investigaciones.

En cuanto al ámbito legal, para poder comprobar si las herramientas de saliencia se aproximan cualitativamente a la realidad, será necesario realizar experimentos y siempre que se realizan experimentos con sujetos humanos se deben seguir una serie de principios éticos y legislativos como la Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal (LOPD) y el Real Decreto 1720/2007, de 21 de Diciembre (RDLOPD).

## 5.3 Presupuesto

COSTOS DIRECTOS PERSONAL				
Apellidos y nombre	Puesto	Dedicación (horas)	Precio Unitario (€/hora)	Total (€)
Aramendi Campón Diego	Ingeniero Junior	250	40	10.000
Peláez Moreno Carmen	Ingeniero Jefe	30	120	3.600
<b>Total</b>				<b>13.600</b>

COSTOS SOFTWARE, EQUIPO, OTROS					
Descripción	Coste (Con I.V.A.)	Uso (%)	Dedicación (meses)	Depreciación (meses)	Coste por Amortización (€)
Ordenador ASUS A55V	800	100	8,5	50	425
Licencia Matlab	2000	100	8,5	166,67	1.416,7
Licencia IEE Xplore	155	100	8,5	12,92	109,82
<b>Total</b>					<b>1.951,73</b>



<b>COSTOS TOTALES</b>	<b>€</b>
Costos Directos Personal	13.600
Costos por desplazamiento y otros	468
Amortizaciones	1.951,73
Total sin I.V.A.	14.019,73
I.V.A. (18%)	2.523,55
<b>Total</b>	<b>18.543,28</b>

**Tabla 11 Presupuesto**

El precio total del proyecto de investigación asciende a DIECIOCHO MIL QUINIENTOS CUARENTA Y TRES EUROS CON VEINTIOCHO CÉNTIMOS DE EURO

# Capítulo 6

## Conclusiones y Líneas Futuras

### 6.1 Conclusiones

Este proyecto nace con el objetivo principal de adaptar una herramienta de detección de saliencia en imágenes estáticas a la detección de saliencia dinámica o saliencia en secuencias de vídeo. Otro objetivo ha sido comprender y analizar los diferentes estudios e investigaciones, es decir realizar un amplio estudio del estado del arte para obtener conclusiones de la adaptación desarrollada y obtener posibles líneas futuras de trabajo.

En primer lugar se analizó el algoritmo de segmentación utilizado por la herramienta para decidir qué valores tendrían los parámetros que se utilizarían para una correcta segmentación de los objetos de cada plano como se expuso en el Capítulo 3, sección 3.2.

Una vez realizado este análisis, se comprueba que existe cierta inestabilidad entre planos consecutivos cuando no existen cambios bruscos en los píxeles de uno y otro, por lo que posteriormente se proponen diferentes ideas que nos muestran otro problema nuevo, el cambio en el etiquetado de los segmentos de un plano a otro. Al final ninguna las primeras soluciones propuestas, como el seguimiento de los objetos o segmentos dan un buen

resultado, por lo que se decide cambiar de rumbo y se lleva a cabo la idea de realizar un filtrado espacial para intentar suavizar estos cambios bruscos.

Pero los resultados no son demasiado favorecedores todavía, aunque sí se puede observar que se consolidan enlaces entre pequeños segmentos con segmentos de su alrededor por los que son absorbidos. Además se percibe que desaparece el problema de los cambios en el etiquetado de los segmentos.

Para continuar mejorando la salida del segmentador y convertirla en más estable, se lleva a cabo un proceso de filtrado temporal sobre el filtrado espacial ya realizado, tras el cual los resultados en la segmentación producen una salida más estable.

Tras estos resultados podemos concluir que el trabajo realizado no es suficiente para generar una buena estabilidad en la saliencia, principalmente porque necesitaríamos información sobre el movimiento que explicamos en las líneas futuras, además de proponer algunas otras ideas a llevar a cabo con lo que podría mejorar en conjunto del trabajo realizado.

## 6.2 Conclusions

This project was created with the main purpose of adjust a static saliency detection tool to create a dynamic saliency detection tool in video image sequences. Another objective was to understand and analyze the different studies and researches, that is, a comprehensive study of the state of the art and obtain conclusions of the adjust developed and reflect possible future work.

First, we did an analysis for the segmentation algorithm used by the tool to decide what parameters values should be used for the best segmentation of the objects in each frame as discussed in Chapter 3, Section 3.2.

After this analysis, it is found that there is some instability between consecutive planes when there are no abrupt changes in the pixels of each other, then different ideas showed us to propose another new problem, the change in the label segments from one plane to another. At the end any proposed solutions, such as tracking of objects or segments give a good result, so decide to alter the course and holds the idea of spatial filtering to try soft the abrupt changes.

But the results are not good enough, it is possible to note links between small segments with segments around those who are absorbed consolidated. Also perceived that the problem of changes in the labeling of segments disappears.

To further improve the output of the segmentation tool and make it more stable, it performs a temporal filtering process on the spatial filter already performed, after which results in segmentation produce a more stable output.

After these results we can conclude that the work done is not enough to generate a good stability in salience, mainly because we would need information on the movement which is explained in the future work lines where we propose some other ideas to carry out which could be improved the work done.

## 6.3 Líneas futuras

Es evidente que la estabilidad en la saliencia para vídeo que hemos logrado con nuestras propuestas es muy limitada, por lo que se puede mejorar en gran medida y buscar otros desarrollos, añadirle nuevos módulos que permitan lograr los objetivos bien relacionados con otro tipo de filtrados o bien, pensando en otras soluciones alternativas.

En cuanto a mejorar el desarrollo, sería bueno poder detectar automáticamente a partir de qué tamaño resulta un objeto de interés para decidir si analizar su saliencia o desestimarlo considerándolo fondo, esto siempre va a resultar complejo puesto que tenemos diferentes tamaños de imagen y el foco de interés puede estar situado más o menos lejos lo que dificulta en gran medida aportar una solución al problema.

En cuanto a filtros, podemos probar otro tipo de suavizados para estudiar si mejoran la estabilidad, como pudieran ser filtros de mediana, filtros de promedio, filtros Gaussianos u otro tipo de filtros paso bajo derivados de estos para eliminar el ruido. Si trabajamos en el dominio frecuencial también podremos utilizar filtrados paso bajo.

En cuanto a otros módulos resultaría interesante para mejorar la saliencia en vídeo tener en cuenta la trayectoria de los movimientos, qué sucede con los objetos suponiendo una imagen de plano fijo, analizando las características de aceleración y velocidad de los mismos, ya que los descriptores importantes en la saliencia de secuencias de vídeo son todos

aquellos relacionados con el movimiento. En un principio se pensó que se podría llegar a consolidar una aplicación que tuviese en cuenta estas características por lo que en la fase de estudio del arte se tuvo en cuenta la búsqueda de una herramienta para la obtención del flujo óptico.

Otra línea de investigación interesante para la saliencia sería poder contrastar los resultados del algoritmo con datos reales para mejorar la concordancia con el sistema visual humano, para ello deberíamos obtener datos de eye-tracking, para lo que necesitaremos tanto un software como un hardware específico de seguimiento del movimiento ocular cuando el sujeto en cuestión observe el vídeo sobre una pantalla. Siempre tendremos que realizar pruebas al mayor número de sujetos posibles puesto que tanto el hardware como el software de seguimiento actual tiene un elevado margen de error y no alcanza muchas veces el nivel de detalle deseado, por lo que repetiremos el ensayo con muchos sujetos para después calcular algún estadístico sobre los resultados obtenidos. Posteriormente podremos generar un algoritmo de entrenamiento sobre nuestra máquina que nos servirá para mejorar los resultados siempre y cuando se trate de situaciones de vídeo semejantes.

Como se mencionó en el Capítulo 3, sección 3.2 el tipo de secuencias de imágenes es un factor muy importante a la hora de calcular la saliencia, puesto que no podemos entrenar algoritmos para todo tipo de situaciones ni responder correctamente a las mismas, solamente para un conjunto limitado. Es muy importante desarrollar un mecanismo capaz de determinar de forma automática qué tipo de situación se está produciendo en una determinada secuencia de vídeo y entrenar el algoritmo en ese momento para dicha situación o seleccionar el algoritmo previamente entrenado para ello.

# Capítulo 7

## Bibliografía

### 7.1 Referencias bibliográficas

1. Watson, A. B. & Ahumada, A. J.: ‘Model of human visual-motion sensing’. *Optical Society of America, Journal, A: Optics and Image Science* 2, 322–342, 1985
2. Itti, L., Koch, C. and Niebur, E.: ‘A model of saliency-based visual attention for rapid scene analyses. *In pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, no. 11, pp. 1254-1259, 1998.
3. Harel J., Koch, C., and Perona, P.: ‘Graph-based visual saliency, in Advances in Neural Information Processing Systems’. *Proceedings of the Twentieth Annual Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, December 4-7, 2006*, B. Schlkopf, J. C. Platt, and T. Homan, Eds. MIT Press, 2006, pp. 545-552.

4. Hou, X & Zhang, L.: ‘Saliency detection: A spectral residual approach’. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR07)*. IEEE Computer Society, 2007, pp. 1-8.
5. Muratov, O., Zontone, P., Boato, G. and De Natale, F. G. B.: ‘A segment-based image saliency detection’. *In Proc. Of ICASSP2011*, pp.1217–1220 (2011). Doi: 10.1109 /ICASSP. 2011. 5946629.
6. Christoudias, C. M., Georgescu, B. and Meer, P.: ‘Synergism in low level vision’ *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, vol. 4. IEEE, 2002, pp. 150-155.
7. Bagon, S. Boiman O. and Irani M.: ‘What is a Good Image Segment? An unified Approach to Segment Extraction’. *Weizmann Institute of Science, Rehovot, Israel*.
8. Shi, J. & Malik, J.: ‘Normalized cuts and image segmentation’. *PAMI (2000)*
9. Comaniciu, D. & Meer, P.: ‘Mean shift: a robust approach toward feature space analyses. *PAMI (2002)*.
10. Malik, J., Belongie, S., Shi, J., and Leung, T.K.: ‘Textons, contours and regions: Cue integration in image segmentation’. *In: ICCV. (1999)*
11. Galun, M., Sharon, E., Basri, R., and Brandt, A.: ‘Texture segmentation by multiscale aggregation of filter responses and shape elements’. *In: ICCV. (2003)*
12. Kadir, T. & Brady, M.: ‘Unsupervised non-parametric region segmentation using level sets’. *In: ICCV. (2003)*
13. Li, Y., Sun, J., Tang, C.K. and Shum, H.Y.: ‘Lazy snapping’. *ACM TOG (2004)*
14. Rother, C., Kolmogorov, V. and Blake, A.: ‘Interactive foreground extraction using iterated graph cuts’. *In: SIGGRAPH. (2004)*
15. Riklin-Raviv, T., Kiryati, N. and Sochen, N.: ‘Segmentation by level sets and symmetry’. *In: CVPR. (2006)*

16. Borenstein, E. & Ullman, S.: 'Class-specific, top-down segmentation'. *In: ECCV'02*
17. Leibe, B. & Schiele, B.: 'Interleaved object categorization and segmentation'. *In: BMVC'03*
18. Levin, A. & Weiss, Y.: 'Learning to combine bottom-up and top-down segmentation'. *In: ECCV. (2006)*
19. Rother, C., Minka, T., Blake, A. and Kolmogorov, V.: 'Cosegmentation of image pairs by histogram matching - incorporating a global constraint into mrfs'. *In: CVPR'06*
20. Tu, Z., Chen, X., Yuille, A.L. and Zhu, S.C.: 'Image parsing: Unifying segmentation, detection, and recognition'. *IJCV (2005)*.
21. Liu, H., Qiu, X., Huang, Q, Jiang, S. and Xu, C.: 'Advertise gently- inimage advertising with low intrusiveness'. *In Image Processing (ICIP), 2009 16th IEEE International Conference on. IEEE, 2009, pp. 3105-3108.*
22. Frintrop S., Rome E. and Christensen H.I.: 'Computational Visual Attention Systems and their Cognitive Foundation: A Survey'. *ACM Transactions on Applied Perception (TAP) (2010)*
23. Comanicu D. & Meer P.: 'Mean shift: A robust approach toward feature space analysis.' *IEEE Trans. Pattern Anal. Machine Intell., 24, 603-619, May 2002.*
24. Meer P. & Georgescu B.: 'Edge detection with embedded confidence.' *IEEE Trans. Pattern Anal. Machine Intell., 23, 1351-1365, December 2001.*
25. González, R., Woods R, and Eddins S.: '*Digital image processing using MATLAB*'. Pearson Education India, 2004.



26. Liu, C.: *'Beyond Pixels: Exploring New Representations and Applications for Motion Analysis'*. *Doctoral Thesis*. Massachusetts Institute of Technology. May 2009.