

UNIVERSIDAD AUTONOMA DE MADRID

ESCUELA POLITÉCNICA SUPERIOR



**COMBINACIÓN DE EXTRACTORES Y
DESCRIPTORES DE PUNTOS Y REGIONES DE
INTERÉS**

-PROYECTO FIN DE CARRERA-

Fulgencio Navarro Fajardo

Octubre 2013

COMBINACIÓN DE EXTRACTORES Y DESCRIPTORES DE PUNTOS Y REGIONES DE INTERÉS

AUTOR: Fulgencio Navarro Fajardo
TUTOR: Marcos Escudero Viñolo
PONENTE: Jesús Bescós Cano

Grupo VPULab
Dpto. de Tecnología Electrónica y de las Comunicaciones
Escuela Politécnica Superior
Universidad Autónoma de Madrid
Octubre 2013

Palabras clave

Puntos y regiones de interés, SIFT, *superpíxeles*, especificidad de objetos, oclusión, reconocimiento de logos.

Resumen

El objetivo de este proyecto ha sido el de desarrollar un método de descripción de imágenes basado en puntos y regiones de interés a partir del estudio, análisis y detección de defectos de aplicación del principal método del estado del arte de este campo, los puntos SIFT.

El algoritmo de SIFT incorpora información del entorno para llevar a cabo una descripción de la imagen robusta y distintiva. Sin embargo, no lleva a cabo discriminación alguna sobre la procedencia de la información de su entorno que incorpora a dicha descripción.

En aplicaciones en las que la información que se desea extraer de la imagen se encuentra en el objeto, incorporar a la descripción de éste datos del entorno resultaría perjudicial para la aplicación. Especialmente si los fondos sobre los que aparece éste son cambiantes o con mucha información.

El algoritmo desarrollado ha surgido a partir de una combinación del método SIFT con los novedosos segmentadores en regiones conocidos como segmentadores en *Superpíxeles*. Esta combinación ha permitido desarrollar un algoritmo con las mismas y, en algunos casos, superiores propiedades de robustez y distintividad presentadas por SIFT, pero aislando en distintas descripciones la información perteneciente a distintas regiones.

El algoritmo desarrollado se ha evaluado objetivamente para el estudio de su comportamiento con un *toy-example* y una comparativa de propiedades siguiendo el modelo de las realizadas en el estado del arte de este campo.

Por último, se han diseñado dos aplicaciones sobre las que se ha implementado el algoritmo creado. La selección de las aplicaciones ha sido llevada a cabo con la intención de observar el método en una aplicación genérica y en otra más orientada a los defectos de aplicación corregidos. Con este fin, las aplicaciones sobre las que se ha implementado el algoritmo han sido: reconocimiento bajo oclusión y reconocimiento de logotipos en imágenes.

Abstract

The target of this project has been to develop a method for image description based on points and regions of interest. This method has been derived from the study, analysis and detection defects of the application of the main state of the art method in this field, the SIFT points.

The SIFT algorithm incorporates information of the environment to carry out a robust and distinctive description of the image. However, no discrimination is carried out based on the origin of the information of the environment that incorporates this description.

In the applications where the information to be extracted from the image is on an object, incorporating environment's data to its description would be detrimental to the application. Especially if the background where the object is present is changing or they have much information.

The developed algorithm has arisen from a combination of the SIFT method with the newfangled segmentation-based methods known as *Superpixels*. This combination has allowed us to develop an algorithm with the same and even higher robustness and distinctiveness presented by SIFT, but isolating, in the different descriptions the information belonging to different regions

The developed algorithm has been objectively evaluated to study its behavior with a toy-example and a comparison of properties following the model of those made in the state of the art of this field.

Finally, two applications were designed on which the algorithm operation is evaluated. The applications selection has been carried out in order to evaluate the goodness of the techniques designed to correct the application defects. With this in mind, the applications on which the algorithm was evaluated were: recognition under occlusion and logo-recognition in images.

Agradecimientos

Me gustaría agradecer en primer lugar a mi tutor, Marcos Escudero Viñolo, la oportunidad de haber podido llevar a cabo este proyecto. Su ayuda y apoyo al igual que sus directrices han sido piezas claves para que este proyecto llegase a buen puerto.

Del mismo modo, quería agradecer a todos los componentes del VPULab su apoyo y ayuda. El buen ambiente que crean en el lugar de trabajo hace que las cosas sean siempre más sencillas.

No me gustaría olvidarme por supuesto de los jefes del grupo, Jesús Bescós y José María Martínez, pues sin ellos mi proyecto tampoco habría sido posible.

Querría dedicar también unas líneas de estos agradecimientos a mi compañero Juan Manuel. Entre las muchas cosas que me llevo de esta carrera, su amistad y el haber sido su compañero será una de las que más valore.

Mi grupo de amigos y compañeros también tienen cabida aquí, pues de una manera o de otra han sido parte responsable de que este redactando este capítulo.

Pero si hay dos partes realmente culpables de que mi carrera este llegando a su fin, esas son: mi novia Julia, su apoyo todos y cada uno de los días desde que la conocí en esta misma escuela han sido el mejor regalo que he podido recibir aquí; y mis padres y hermano, ellos, que nunca me han fallado, son sin duda alguna la causa principal de que hoy esté donde estoy.

A todos los demás, que por activa o por pasiva habéis formado parte de esta carrera que concluye con este proyecto, con total sinceridad, gracias.

*Fulgencio Navarro Fajardo
Septiembre de 2013*

INDICE DE CONTENIDOS

CAPÍTULO 1 - INTRODUCCIÓN	1
1.1 MOTIVACIÓN	1
1.2 OBJETIVOS.....	2
1.3 ORGANIZACIÓN DE LA MEMORIA	3
CAPÍTULO 2 - ESTADO DEL ARTE	5
2.1 INTRODUCCIÓN.....	5
2.2 PRINCIPALES MÉTODOS PARA LA DESCRIPCIÓN DE IMÁGENES	6
2.2.1 <i>Algoritmos de detección</i>	7
2.2.1.1 Detector de Harris-affine.....	7
2.2.1.2 Detector de Hessian-affine.....	8
2.2.1.3 Detector de SIFT	8
2.2.1.4 Detector de SURF	9
2.2.1.5 Detector de MSER	9
2.2.2 <i>Algoritmos de descripción</i>	10
2.2.2.1 Descriptor de SIFT	10
2.2.2.2 GLOH	10
2.2.2.3 Descriptor basado en contexto de siluetas	11
2.2.2.4 Descriptor de imágenes de spin	11
2.2.2.5 Descriptor de momentos invariantes.....	11
2.3 COMPARATIVA DE MÉTODOS POR PROPIEDADES.....	12
2.3.1 <i>Comparativa de detectores</i>	12
2.3.2 <i>Comparativa de descriptores</i>	13
2.4 CONCLUSIÓN DEL ESTADO DEL ARTE	14
CAPÍTULO 3 – ANÁLISIS DE DETECTOR Y DESCRIPTOR DE SIFT	17
3.1 INTRODUCCIÓN.....	17
3.2 FUNCIONAMIENTO	18
3.2.1 <i>Detección de Extremos en el Espacio Escala</i>	18
3.2.2 <i>Localización de puntos clave estables</i>	19
3.2.3 <i>Asignación de la orientación</i>	20
3.2.4 <i>Descripción del punto de interés</i>	21
3.3 PROPIEDADES.....	22
3.3.1 <i>Invariancia a cambios de escala</i>	22
3.3.2 <i>Invariancia a rotaciones</i>	22
3.3.3 <i>Invariancia a cambios lineales de iluminación</i>	23
3.3.4 <i>Invariancia a borrosidad</i>	23
3.3.5 <i>Invariancia al ruido</i>	24
3.3.6 <i>Invariancia a transformaciones afines</i>	24
3.3.7 <i>Alta distintividad</i>	24
CAPÍTULO 4 – SIFT: DEFECTOS DE APLICACIÓN	25
4.1 INTRODUCCIÓN.....	25
4.2 RECONOCIMIENTO DE OBJETOS E IMÁGENES DUPLICADAS	26
4.3 RECONOCIMIENTO DE OBJETOS ANTE OCLUSIONES	27
CAPÍTULO 5 – ALGORITMO PROPUESTO: REGSIFT	29
5.1 INTRODUCCIÓN.....	29
5.2 ALGORITMO DE DETECCIÓN	31
5.2.1 <i>Detección SIFT</i>	31
5.2.2 <i>Partición en regiones</i>	31
5.3 ALGORITMO DE DESCRIPCIÓN	34
5.3.1 <i>Replicación de features</i>	34

5.3.2	<i>Orientación principal</i>	35
5.3.3	<i>Cálculo del histograma</i>	36
5.3.4	<i>Generación del nuevo descriptor</i>	37
5.4	RESUMEN GRÁFICO DEL ALGORITMO	38
5.5	MÉTRICAS ASOCIADAS	39
5.5.1	<i>Punto a punto</i>	40
5.5.2	<i>Punto a conjunto de puntos</i>	40
5.5.3	<i>Conjunto de puntos a conjunto de puntos</i>	40
CAPÍTULO 6 – EVALUACIÓN DEL ALGORITMO		41
6.1	INTRODUCCIÓN	41
6.2	TOY-EXAMPLE	41
6.2.1	<i>Introducción</i>	41
6.2.2	<i>Metodología de evaluación</i>	41
6.2.2.1	Dataset	41
6.2.2.2	Clases	42
6.2.2.3	Métrica	42
6.2.2.4	Estadísticos	43
6.2.2.5	Objetivo de la evaluación	44
6.2.3	<i>Resultados</i>	44
6.2.4	<i>Análisis de resultados</i>	47
6.3	EVALUACIÓN DE PROPIEDADES	49
6.3.1	<i>Introducción</i>	49
6.3.2	<i>Metodología de evaluación</i>	50
6.3.2.1	Dataset	50
6.3.2.2	Métrica	50
6.3.3	<i>Resultados</i>	50
6.3.3.1	Blur	51
6.3.3.2	Light	53
6.3.3.3	Viewpoint	55
6.3.3.4	Scale + Rotate	57
6.3.4	<i>Análisis de resultados</i>	59
CAPÍTULO 7 – APLICACIONES		63
7.1	INTRODUCCIÓN	63
7.2	RECONOCIMIENTO DE LOGOTIPOS	63
7.2.1	<i>Metodología de evaluación</i>	63
7.2.2.1	Dataset	63
7.2.2.2	Métrica	64
7.2.2	<i>Diseño</i>	64
7.2.3	<i>Resultados</i>	65
7.2.4	<i>Análisis de resultados y evaluación</i>	70
7.3	RECONOCIMIENTO ANTE OCLUSIONES	72
7.3.1	<i>Metodología de evaluación</i>	72
7.3.2.1	Dataset	72
7.3.2.2	Métrica	73
7.3.2	<i>Diseño</i>	74
7.3.3	<i>Resultados</i>	75
7.3.1.1	Resultados generales	76
7.3.1.2	Resultados de región borde de oclusión	77
7.3.4	<i>Análisis de resultados</i>	79
CAPÍTULO 8 – CONCLUSIONES Y TRABAJO FUTURO		81
8.1	CONCLUSIONES	81
8.2	TRABAJO FUTURO	83
REFERENCIAS		87
GLOSARIO		I

ANEXOS	I
A DATASET DEL TOY-EXAMPLE	I
A.1. Fondos Poco texturados.....	I
A.2. Fondos Altamente texturados	I
A.3. Máscaras de ground-truth y cromas	II
B DATASET PARA ANÁLISIS DE PROPIEDADES	III
B.1. Imágenes de Blur	III
B.2. Imágenes de cambio de iluminación	IV
B.3. Imágenes de cambio de punto de vista	V
B.4. Imágenes cambio de escala + rotación.....	VI
C DATASET LOGOS RECOGNITION.....	VII
D DATASET RECONOCIMIENTO ANTE OCLUSIONES	IX
D.1 Secuencia Objeto oclusivo Poco Texturado.....	IX
D.2 Secuencia Objeto oclusivo Altamente Texturado	X
D.3 Secuencia Objeto doble oclusión y propiedades.....	XII
E EVOLUCIÓN DEL ALGORITMO	XV
E.1 Introducción.....	XV
E.2 Línea de evolución	XV
E.3 Desarrollo	XV
E.3.1 Etapa 1: SIFT	XV
E.3.2 Etapa 2: MSER	XVI
E.3.3 Etapa 3: Revisión y Re-selección.....	XVIII
E.3.4 Etapa 4: Superpíxeles	XIX

INDICE DE FIGURAS

FIGURA 1-1: EJEMPLOS DE TAREAS.	1
FIGURA 2-1: MÉTODOS SELECCIONADOS PARA ESTUDIO.	6
FIGURA 2-2: (A) ANÁLISIS GAUSSIANO, (B) NORMALIZACIÓN AFÍN Y (C) PUNTOS RESULTANTES.....	7
FIGURA 2-3: (A) PIRÁMIDES GAUSSIANAS ESPACIO ESCALA Y DOG, (B) ANÁLISIS LOCALIZADO Y (C) RESULTADO DE DETECCIÓN....	8
FIGURA 2-4: (A) ANÁLISIS MULTIESCALA, (B) NORMALIZACIÓN DE ESPACIO Y (C) RESULTADO DE DETECCIÓN.....	9
FIGURA 2-5: MSER DETECTADAS CON DISTINTOS UMBRALES DE ESTABILIDAD.	9
FIGURA 2-6: (A) ORIENTACIÓN PRINCIPAL, (B) GRADIENTES ORIENTADOS Y (C) VECTOR DE DESCRIPCIÓN	10
FIGURA 2-7: MODELO DE ANÁLISIS MULTIESCALA GLOH	10
FIGURA 2-8: (A) FIGURA CON DETECCIONES, (B) PUNTOS SOBRE <i>SPIN IMAGE</i> Y (C) MODO DE INTERPOLACIÓN DE LA INTENSIDAD.11	11
FIGURA 3-1: ESPACIO ESCALA GAUSSIANO Y DOG.....	19
FIGURA 3-2: LOCALIZACIÓN DE MÁXIMOS Y MÍNIMOS LOCALES.....	19
FIGURA 3-3: DESCRIPCIÓN DE LOS PUNTOS DE INTERÉS	22
FIGURA 3-4: SIFT SOBRE CAMBIO DE ESCALA.....	22
FIGURA 3-5: SIFT SOBRE ROTACIÓN	23
FIGURA 3-6: SIFT SOBRE ILUMINACIÓN LINEAL.	23
FIGURA 3-7: SIFT SOBRE BORROSIDAD.	23
FIGURA 3-8: SIFT SOBRE RUIDO.	24
FIGURA 3-9: SIFT SOBRE TRANSFORMACIÓN AFÍN.	24
FIGURA 4-1: ESQUEMA DE APLICACIONES CON SIFT Y PROPIEDADES REQUERIDAS	25
FIGURA 5-1: DETALLE DE DOS <i>FEATURES</i> DE SIFT	29
FIGURA 5-2: COMPARACIÓN DE DESCRIPTORES DE <i>FEATURE1</i> Y <i>FEATURE 2</i>	29
FIGURA 5-3: EJEMPLO DE <i>FEATURES</i> A DESCRIBIR EN <i>REGSIFT</i>	30
FIGURA 5-4: DESCRIPTORES DE (A) <i>FEATURES</i> 1 Y 2 DE SIFT ORIGINAL, (B) <i>FEATURE1'</i> Y <i>FEATURE2'</i> Y (C) <i>FEATURE1''</i> Y <i>FEATURE2''</i> DE <i>REGSIFT</i>	30
FIGURA 5-5: ESQUEMA DE DETECCIÓN DE <i>REGSIFT</i>	31
FIGURA 5-6: EJEMPLO DE SEGMENTACIÓN POR <i>SUPERPÍXELES</i>	32
FIGURA 5-7: EJEMPLO DE <i>SUPERPÍXELES</i> EXTRAÍDOS EN DIFERENTES OCTAVAS.....	33

FIGURA 5-8: ESQUEMA DE DESCRIPCIÓN DE REGSIFT.....	34
FIGURA 5-9: EJEMPLO DE REPLICACIONES POR ÁREA DE REGIÓN Y MÁSCARAS GENERADAS	35
FIGURA 5-10: DETALLE DE ORIENTACIÓN DEL DESCRIPTOR. A) DESCRIPTOR SIFT B),C) Y D) DESCRIPTORES <i>REGSIFT</i>	36
FIGURA 5-11: ESQUEMA DE DESCRIPTORES INDIVIDUALES DE <i>FEATURES REGSIFT</i>	37
FIGURA 5-12: ESQUEMA DE CONFORMACIÓN DE DESCRIPTOR DE PUNTO SIFT POR <i>REGSIFT</i>	38
FIGURA 5-13: IMAGEN DE ANÁLISIS Y DETALLE DE DETECCIÓN SIFT CON ÁREA DE DESCRIPCIÓN SUPERPUESTA.	38
FIGURA 5-14: DETALLE DEL RESULTADO DE LA SEGMENTACIÓN EN <i>SUPERPÍXELES</i>	38
FIGURA 5-15: ÁREAS DE DESCRIPTORES DE BORDES DIVIDAS POR REGIONES.	39
FIGURA 5-16: OBTENCIÓN DE LOS VECTORES POR REGIONES, <i>REGSIFT</i> DESCRIPTOR, O SU CONJUNCIÓN <i>GROUPEDREGSIFT</i> DESCRIPTOR.	39
FIGURA 5-17: RELACIÓN INYECTIVA.	40
FIGURA 6-1: I01 E IT01	42
FIGURA 6-2: TP / FP PARA FONDOS POCO TEXTURADOS.....	44
FIGURA 6-3: PRECISIÓN/RECALL PARA FONDOS POCO TEXTURADOS.	45
FIGURA 6-4: FSCORE PARA FONDOS POCO TEXTURADOS.....	45
FIGURA 6-5: TP / FP PARA FONDOS ALTAMENTE TEXTURADOS.	46
FIGURA 6-6: PRECISIÓN/RECALL PARA FONDOS ALTAMENTE TEXTURADOS.....	46
FIGURA 6-7: FSCORE PARA FONDOS ALTAMENTE TEXTURADOS.	47
FIGURA 6-8: EFECTO DE SEPARACIÓN DE GAUSSIANAS.	48
FIGURA 6-9: EJEMPLOS DEL <i>DATASET</i> POR FAMILIAS DE PROPIEDADES.....	50
FIGURA 6-10: <i>DATASET</i> DE IMÁGENES DE <i>BLUR</i> . DE MENOR A MAYOR <i>BLUR</i> (IZDA. A DCHA. Y DE ARRIBA A ABAJO)	51
FIGURA 6-11: FDP DE DISTANCIAS PARA <i>BLUR01-BLUR02</i>	51
FIGURA 6-12: FDP DE DISTANCIAS PARA <i>BLUR01-BLUR03</i>	52
FIGURA 6-13: FDP DE DISTANCIAS PARA <i>BLUR01-BLUR04</i>	52
FIGURA 6-14: FDP DE DISTANCIAS PARA <i>BLUR01-BLUR06</i>	52
FIGURA 6-15: FDP DE DISTANCIAS PARA <i>BLUR01-BLUR06</i>	53
FIGURA 6-16: <i>DATASET</i> DE IMÁGENES DE <i>LIGHT</i> . DE MENOR A MAYOR <i>LIGHT</i> (IZDA. A DCHA. Y DE ARRIBA A ABAJO)	53
FIGURA 6-17: FDP DE DISTANCIAS PARA <i>LIGHT01- LIGHT 02</i>	53
FIGURA 6-18: FDP DE DISTANCIAS PARA <i>LIGHT01- LIGHT 03</i>	54
FIGURA 6-19: FDP DE DISTANCIAS PARA <i>LIGHT01- LIGHT 04</i>	54
FIGURA 6-20: FDP DE DISTANCIAS PARA <i>LIGHT01- LIGHT 05</i>	54
FIGURA 6-21: FDP DE DISTANCIAS PARA <i>LIGHT01- LIGHT 06</i>	55
FIGURA 6-22: <i>DATASET</i> DE IMÁGENES DE <i>VIEWPOINT</i> . DE MENOR A MAYOR ÁNGULO DE <i>VIEWPOINT</i> (IZDA. A DCHA. Y DE ARRIBA A ABAJO).....	55
FIGURA 6-23: FDP DE DISTANCIAS PARA <i>VIEWPOINT01- VIEWPOINT02</i>	55
FIGURA 6-24: FDP DE DISTANCIAS PARA <i>VIEWPOINT01- VIEWPOINT03</i>	56
FIGURA 6-25: FDP DE DISTANCIAS PARA <i>VIEWPOINT01- VIEWPOINT04</i>	56
FIGURA 6-26: FDP DE DISTANCIAS PARA <i>VIEWPOINT01- VIEWPOINT05</i>	56
FIGURA 6-27: FDP DE DISTANCIAS PARA <i>VIEWPOINT01- VIEWPOINT06</i>	57
FIGURA 6-28: <i>DATASET</i> DE IMÁGENES DE <i>SCALE+ROTATE</i>	57
FIGURA 6-29: FDP DE DISTANCIAS PARA <i>SCL+ROT01- SCL+ROT02</i>	57
FIGURA 6-30: FDP DE DISTANCIAS PARA <i>SCL+ROT01- SCL+ROT03</i>	58
FIGURA 6-31: FDP DE DISTANCIAS PARA <i>SCL+ROT01- SCL+ROT04</i>	58
FIGURA 6-32: FDP DE DISTANCIAS PARA <i>SCL+ROT01- SCL+ROT05</i>	58
FIGURA 6-33: FDP DE DISTANCIAS PARA <i>SCL+ROT01- SCL+ROT06</i>	59
FIGURA 7-1: IMÁGENES POR FAMILIAS DE LOGOTIPOS. A)CANAL+HD B)MTV-HD C)EUROSPORT D)LASIXTA HD E)TELECINCO HD	64
FIGURA 7-2: ESQUEMA DE DISEÑO DE LA APLICACIÓN.	65
FIGURA 7-3: IMÁGENES POR CATEGORÍAS. A) OCLUIDO POCO TEXTURADO B) OCLUIDO ALTAMENTE TEXTURADO C) EVALUACIÓN DE PROPIEDADES C) OCLUIDO DOS ELEMENTO	72
FIGURA 7-4: <i>GROUND-TRUTH</i> POR CATEGORÍAS. A) OCLUIDO POCO TEXTURADO B) OCLUIDO ALTAMENTE TEXTURADO C) EVALUACIÓN DE PROPIEDADES C) OCLUIDO DOS ELEMENTO	73
FIGURA 7-5: ESQUEMA DE ETAPAS DE APLICACIÓN DE OCLUSIONES.	74
FIGURA 7-6: VALORES DE PRECISIÓN GENERALES POR CATEGORÍA Y MÉTODO.	76
FIGURA 7-7: VALORES DE <i>RECALL</i> GENERALES POR CATEGORÍA Y MÉTODO.	76
FIGURA 7-8: VALORES DE <i>FSCORE</i> GENERALES POR CATEGORÍA Y MÉTODO.	76
FIGURA 7-9: DETALLE DE REGIÓN BORDE DE OCLUSIÓN.	77

FIGURA 7-10: VALORES DE PRECISION EN BORDE DE OCLUSION POR CATEGORÍA Y METODO.	78
FIGURA 7-11: VALORES DE <i>RECALL</i> EN BORDE DE OCLUSION POR CATEGORÍA Y METODO.	78
FIGURA 7-12: VALORES DE <i>FSCORE</i> EN BORDE DE OCLUSION POR CATEGORÍA Y METODO.	78

INDICE DE TABLAS

TABLA 2-1: COMPARATIVA DE DETECTORES.....	13
TABLA 2-2: COMPARATIVA DE DESCRIPTORES	14
TABLA 6-1: CLASES DE PUNTOS <i>TOY-EXAMPLE</i>	42
TABLA 6-2: MÉTRICA <i>REGSIFT TOY-EXAMPLE</i>	43
TABLA 6-3: MÉTRICA <i>SIFT TOY-EXAMPLE</i>	43
TABLA 7-1: RESULTADOS PARA REPRESENTANTE DE <i>FAMILIA1</i>	66
TABLA 7-2: RESULTADOS PARA REPRESENTANTE DE <i>FAMILIA2</i>	67
TABLA 7-3: RESULTADOS PARA REPRESENTANTE DE <i>FAMILIA3</i>	68
TABLA 7-4: RESULTADOS PARA REPRESENTANTE DE <i>FAMILIA4</i>	69
TABLA 7-5: RESULTADOS PARA REPRESENTANTE DE <i>FAMILIA5</i>	70
TABLA 7-6: VALORES MEDIOS DE PRECISIÓN, <i>RECALL</i> Y <i>FSCORE</i> GENERALES APLICACIÓN DE OCLUSIONES.	77
TABLA 7-7: PORCENTAJES DE MEJORA <i>FSCORE</i> EN APLICACIÓN DE OCLUSIONES	77
TABLA 7-8: VALORES MEDIOS DE PRECISIÓN, <i>RECALL</i> Y <i>FSCORE</i> EN BORDE DE OCLUSION, APLICACION DE OCLUSIONES.	79
TABLA 7-9: PORCENTAJES DE MEJORA <i>FSCORE</i> EN APLICACIÓN DE OCLUSIONES	79

INDICE DE ECUACIONES

ECUACIÓN 1.....	8
ECUACIÓN 2.....	11
ECUACIÓN 3.....	18
ECUACIÓN 4.....	18
ECUACIÓN 5.....	20
ECUACIÓN 6.....	20
ECUACIÓN 7.....	20
ECUACIÓN 8.....	20
ECUACIÓN 9.....	35
ECUACIÓN 10.....	40
ECUACIÓN 11.....	40
ECUACIÓN 12.....	40
ECUACIÓN 13.....	40
ECUACIÓN 14.....	40
ECUACIÓN 15.....	40
ECUACIÓN 16.....	43
ECUACIÓN 17.....	43
ECUACIÓN 18.....	43

Capítulo 1 - Introducción

1.1 Motivación

La extracción automática de información de interés en imágenes digitales constituye hoy en día, y a pesar de los esfuerzos realizados en la materia, uno de los principales retos del campo del tratamiento de imagen. Su correcta extracción es la base de tareas de alto nivel como la detección y reconocimiento de objetos [1][2][3], la reconstrucción de escenarios 3D [4] o el modelado de secuencias de vídeo[5].

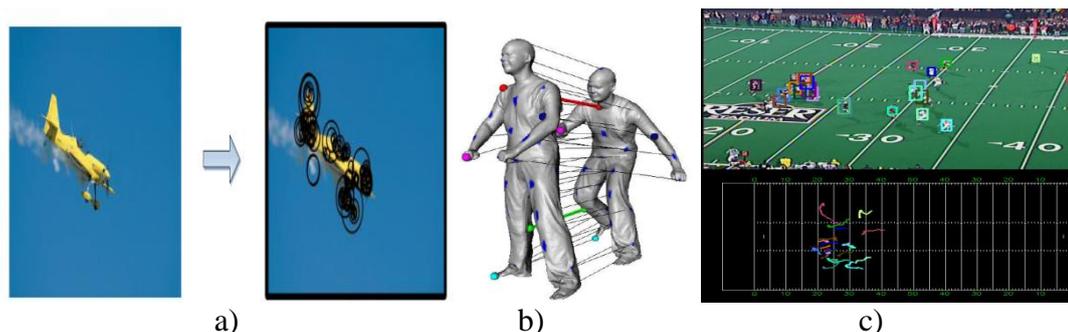


Figura 1-1: Ejemplos de tareas. Fuente: a):[6] b):[7] c):[8]

Si bien se ha dedicado un amplio esfuerzo por parte de la comunidad científica al diseño de la aplicación final [9][10][11][12], las bases de extracción y descripción han sido asumidas como correctas y raramente replanteadas. Este proyecto busca entrar en el fondo de la técnica, haciendo posible la corrección de posibles errores del método en aplicaciones para las que no fueron diseñados, pero para las que actualmente se utilizan.

Entre todos los métodos de extracción y descripción en imágenes que se encuentran en la actualidad en el estado del arte, los puntos y las regiones presentan una mayor relevancia [13].

Si se utilizase como criterio el número de referencias (#20053 fuente: *Google Scholar*), los puntos **SIFT** [1] son, sin lugar a duda, el método más empleado. Sin embargo, como la mayoría de métodos en el área del tratamiento de imágenes, éste surgió como respuesta a una aplicación concreta, y al generalizarse a otras presenta deficiencias que en su aplicación no resultaban ser tales.

En particular, los puntos SIFT surgieron para la correspondencia y *matching* entre imágenes estáticas, esto es, extraían información relevante de toda una imagen y buscaban encontrarla en el mismo escenario pero bajo diferentes condiciones de captura (cambios de iluminación, cambios de punto de vista, cambios de escala, etc.). En consecuencia, su uso en otras aplicaciones tales como el reconocimiento de un objeto en escenarios de “diferente naturaleza”, e.g. variando el fondo, donde solo dicho objeto presentará información relevante, generará defectos derivados de la incorporación de información del entorno en el que se encuentra.

El hecho de incluir en la descripción de un objeto información del entorno puede dar lugar a dos situaciones con consecuencias dispares. Si la detección tiene lugar en el interior del objeto, la información del entorno resulta positiva. Si la detección tiene lugar en el borde del objeto, se incluirá información del fondo en la descripción del objeto, lo cual

resultará muy negativo, principalmente en casos de entornos muy cambiantes o con mucha información.

Pese a que, como podría parecer, una situación contrarresta a la otra, la mayoría de detecciones de SIFT tienen lugar en los bordes, por lo que la segunda situación resulta mucho más relevante.

Esta situación resulta por lo tanto un grave problema para el reconocimiento de objetos mediante puntos SIFT. Aun así, es cierto que los puntos SIFT siguen mejorando el comportamiento de la gran mayoría de métodos [14].

Como resultado de estas reflexiones la motivación principal de este proyecto será adaptar el comportamiento de una herramienta específica para su uso en otras aplicaciones distintas de la originalmente planteada. En definitiva, se buscará mejorar los resultados del estado del arte actual de dicha herramienta en aplicaciones de las citadas características.

1.2 Objetivos

Establecida la motivación de este proyecto, se afronta el objetivo fundamental del mismo, que será el desarrollo de un método de extracción y descripción automática de características en imágenes que cumpla las siguientes premisas:

- Uso generalizable de los descriptores.
- Estabilidad en un ámbito genérico de análisis.
- Mejora de las deficiencias de los métodos en el estado del arte en aplicaciones concretas.

Para abordar este objetivo, se enfocará el trabajo desde una serie de objetivos parciales que se supone permitirán alcanzar el objetivo global de una manera progresiva y adecuada. Estos objetivos desglosados serán:

- Estudio detallado del estado del arte:
Se analizarán los distintos métodos de descripción de imágenes mediante puntos o regiones de interés (PROIID) del estado del arte actual y se compararán las propiedades que presentan bajo diversas condiciones
- Análisis en profundidad del método seleccionado:
Una vez seleccionado el método o métodos con mejores propiedades[14][15], se estudiará en detalle su funcionamiento y comportamiento.
- Estudio crítico de aplicaciones:
Se evaluarán las distintas aplicaciones en las que dicho método haya sido empleado en busca de aplicaciones en las que se requieran unas propiedades no presentes en su diseño. De este punto saldrá el objetivo concreto del algoritmo a desarrollar.
- Desarrollo de algoritmo:
De acuerdo con las propiedades o comportamientos a mejorar, se desarrollará un algoritmo que, alterando en la menor medida posible las propiedades del método seleccionado, aporte nuevas capacidades mediante

la inclusión de herramientas aplicadas en su proceso. Se incluirá un apartado de métrica básica de evaluación asociada al algoritmo que será la que se use durante su evaluación.

- Evaluación de propiedades:
De manera similar a [14][15][16], se evaluarán las nuevas propiedades del método implementado para medir cuantitativamente la calidad del método diseñado.
- Evaluación sobre aplicaciones:
Se desarrollarán dos aplicaciones sobre las que probar los comportamientos, tanto del método original como del desarrollado, con el fin de comprobar si se ha conseguido solventar el problema detectado y mejorar los resultados obtenidos.

Así pues, se puede concluir que, el objetivo de este proyecto no es otro sino el de mejorar el estado del arte de la PRoIID, y más específicamente el del método seleccionado sobre aplicaciones no contempladas en su diseño original.

1.3 Organización de la memoria

La memoria del proyecto va a constar de los siguientes capítulos:

- **Capítulo 1:** Introducción y motivación del proyecto.
- **Capítulo 2:** Estado del arte de PRoIID. Estudio de los principales métodos de detección y descripción de imágenes. Análisis comparativo de propiedades y elección de los métodos a mejorar (SIFT).
- **Capítulo 3:** Desarrollo en profundidad del método SIFT, funcionamiento, propiedades y aplicaciones.
- **Capítulo 4:** Detección de defectos asociados a la aplicación de SIFT.
- **Capítulo 5:** Descripción del método desarrollado. Herramientas, técnicas y algoritmos utilizados para lograr paliar los defectos de aplicación hallados en el capítulo 4. Métrica recomendada.
- **Capítulo 6:** Análisis comparativo en términos cualitativos y cuantitativos de resultados entre los métodos del estado del arte y el propio. *toy-example* y evaluación de propiedades.
- **Capítulo 7:** Exposición de aplicaciones reales desarrolladas para la evaluación del algoritmo implementado. Reconocimiento de objetos ocluidos y reconocimiento de logotipos en diversos ámbitos.
- **Capítulo 8:** Conclusiones obtenidas tras el análisis de resultados. Problemas pendientes y líneas futuras de investigación abiertas.
- **Referencias, glosario y anexos.**

Capítulo 2 - Estado del arte

2.1 Introducción

En la actualidad se manejan volúmenes de información muy grandes. Gran parte de esta información se encuentra almacenada en forma de imágenes y vídeos, esto es, de forma visual. Si bien es cierto que el almacenamiento de dicha información ha alcanzado niveles de automatismo de casi un 100%, su procesado sigue siendo aún una tarea llevada a cabo de forma supervisada.

La principal causa de este hecho es que los distintos segmentos de información (píxeles) almacenados sobre una misma imagen o secuencia de imágenes carecen de sentido en sí mismos. El conjunto de ellos presentan un sentido que requiere una interpretación subjetiva de lo que simbolizan, pero de manera individual carecen de identidad semántica.

Pese a que los sistemas que trabajan sobre imágenes completas han alcanzado resultados relativamente satisfactorios para tareas complejas, el nivel subjetivo derivado de la infinidad de formas, posiciones, colores y demás características que pueden formar una escena, hace que el desarrollo de los mismos quede limitado por el funcionamiento de los algoritmos de más bajo nivel del propio sistema.

Habida cuenta del problema citado en el párrafo anterior, se tuvo que dar un nuevo enfoque a la forma de tratar la información contenida en una imagen. Así, surgieron los descriptores de imagen. Su objetivo no fue otro sino el de describir una imagen basándose en puntos, regiones o agrupaciones de interés.

Según los principios de entropía e información, las variaciones en un entorno estable presentarán mucha mayor información que el resto de elementos invariantes. De igual manera, se definirá así el concepto de elemento de interés en una imagen como aquellos puntos o regiones que presenten una mayor cantidad de información que el entorno en el que se encuentren.

Dentro del conjunto de los descriptores de imagen, se ha de trabajar sobre dos grandes funcionalidades con el fin de lograr una serie de propiedades deseadas. Estas funcionalidades son: *detección* y *descripción*.

Detección:

Consiste en la localización de puntos o áreas de la imagen que presentan la información que se desea describir. En muchas ocasiones, el elemento de detección va ligado con el tipo de descripción que se va a llevar a cabo. Así pues, se pueden encontrar detectores de esquinas [13], de regiones [17] y de bordes [18]. Por otra parte, la técnica de detección también varía de unos métodos a otros, pudiéndose encontrar descriptores que trabajen con los píxeles de la propia imagen, con sus histogramas o con sus gradientes. Sin embargo, adicionalmente a otras propiedades que se pueden lograr o que pueden ser requeridas por la futura descripción, y que se analizarán más adelante en cada caso, todas las detecciones persiguen un mismo fin: la **repetibilidad**. Todas las detecciones buscan ser capaces de identificar un mismo elemento de interés presente en distintos entornos, puntos de vista y condiciones de imagen.

Descripción:

Para que la descripción de los elementos de interés de una imagen sea la deseada, resulta necesario que estos sean detectados de manera precisa y correcta. Ésta busca aportar nuevas propiedades a estos elementos que permitan, entre otros, corregir posibles defectos de la detección. Se puede decir que, la descripción busca caracterizar los distintos elementos detectados en una imagen de tal manera que estos queden representados de una forma en la que se logren propiedades de estabilidad, capacidad distintiva y de nuevo, **repetibilidad**, que faciliten el objetivo para el que son detectados.

Así pues, y tal y como se comentó en la sección 1.2, el objetivo principal de este proyecto es el de modificar estas dos etapas con el fin de mejorar el funcionamiento de los métodos del estado del arte de la descripción de imágenes.

Este capítulo resume los principales detectores y descriptores existentes en el estado del arte, y se organiza de la siguiente forma:

- i) Estudio de los principales métodos del estado del arte en descripción de imágenes, diferenciando entre las fases de detección y descripción.
- ii) Análisis comparativo de las propiedades de los distintos métodos descritos.
- iii) Estudio pormenorizado de los detectores y descriptores seleccionados para sus posibles mejoras. Análisis de defectos en función de su aplicación.
- iv) Conclusiones y, derivando de estas, selección de propiedades objetivo de mejora en función de la aplicaciones final.

2.2 Principales métodos para la descripción de imágenes

No resulta objetivo de este proyecto un estudio detallado de todos y cada uno de los métodos desarrollados para la descripción de imágenes. Sin embargo, se llevará a cabo un breve análisis y una posterior comparativa de los principales métodos del estado del arte. Para ello, se estudiarán aquellos citados en el artículo comparativo de Mikolajczyk [14], artículo de referencia en el área, donde ya se procedieron a seleccionar los más relevantes y de mejor funcionamiento.

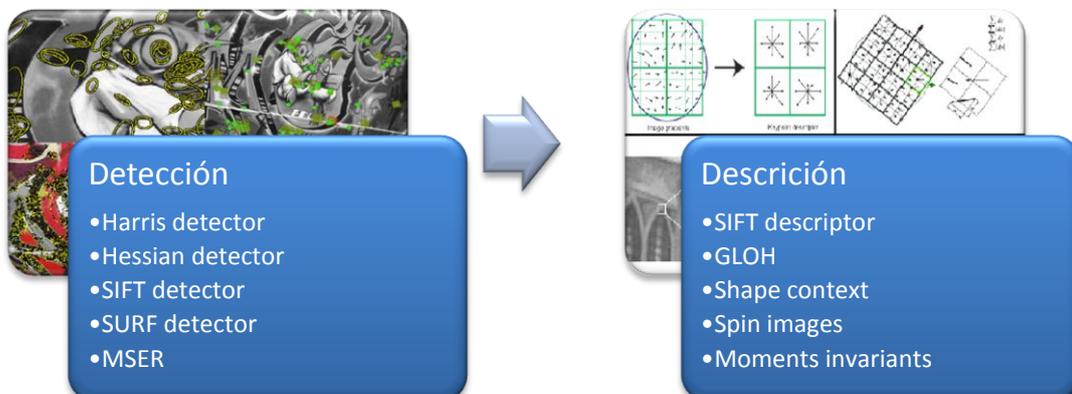


Figura 2-1: Métodos seleccionados para estudio. Fuente: Propia

Una vez determinadas los métodos se procederá de la siguiente manera.

- i) Se expondrán una serie de detectores[13][15][17][18] y descriptores [1][14][19] analizando su funcionamiento a grandes rasgos y las aplicaciones en las que se emplean.
- ii) Se compararán los métodos expuestos desde el puntos de vista de sus comportamientos y propiedades siguiendo el guión de artículos comparativos de referencia [14][15][16].
- iii) Por último, en función de la comparativa realizada, las referencias y aplicaciones en las que es empleado, y las implementaciones disponibles, se seleccionará el método objetivo de mejora.

2.2.1 Algoritmos de detección

2.2.1.1 Detector de Harris-affine

Se trata de un detector que busca ser invariante a cambios de iluminación y a transformaciones afines. Para ello, se basa en el detector de esquinas de Harris [13], un análisis multiescala a través del espacio Laplaciano-Gaussiano y una normalización afín mediante un algoritmo de adaptación de contorno. Llegando así al algoritmo completo expuesto en [15].

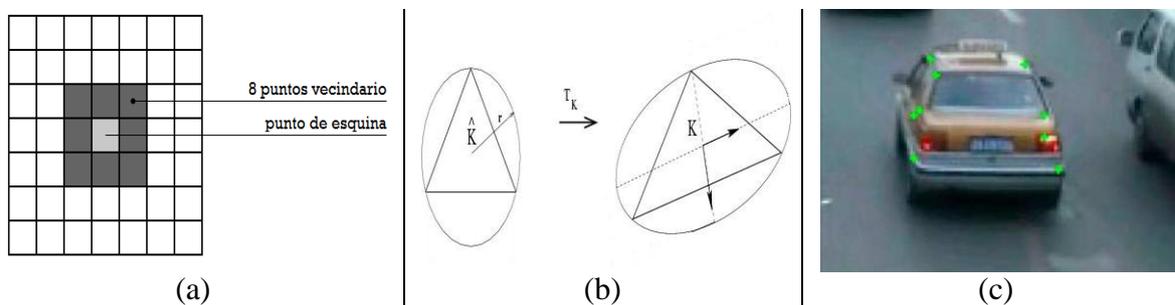


Figura 2-2: (a) Análisis Gaussiano, (b) normalización afín y (c) puntos resultantes. *Fuente:* a) Propia, b) Propia y c)[27]

Este detector es usado principalmente en aplicaciones tales como:

- Reconocimiento y categorización de objetos[20][21].
- Reconocimiento en vídeo [22].
- Reconocimiento de modelos [20][23].

Sus numerosas referencias y sus aplicaciones lo convierten en un referente en este estudio.

2.2.1.2 Detector de Hessian-affine

El detector de Hessian-Affine resulta ser prácticamente idéntico al detector de Harris-Affine, sección 2.2.1.1 . Su principal diferencia radica en que, el detector de Harris-Affine extrae los puntos con el detector de esquinas de Harris[13] basado en la matriz del segundo momento, mientras que el detector de Hessian [15][24] lo hace de la matriz Hessiana:

$$H(f) = \begin{bmatrix} \frac{\delta^2 f}{\delta x_1^2} & \dots & \frac{\delta^2 f}{\delta x_1 x_n} \\ \vdots & \ddots & \vdots \\ \frac{\delta^2 f}{\delta x_n x_1} & \dots & \frac{\delta^2 f}{\delta x_n^2} \end{bmatrix} \quad \text{Ecuación 1}$$

De igual manera, este detector está presente en las mismas aplicaciones que el detector de Harris[21][22][23].

2.2.1.3 Detector de SIFT

El detector de SIFT no es un método en sí mismo, sino que resulta una parte del método de detección y descripción SIFT propuesto por D. Lowe [24]. Los puntos SIFT presentan una elevada cantidad de propiedades debidas en parte a su complejo proceso de detección. La detección es llevada a cabo a través de un método de pirámides Gaussianas de espacio-escala. A partir de sus diferencias (*DOG*) en la propia escala y las contiguas, estudia máximos o mínimos locales y toma el punto central y su vecindario como región detectada. Este comportamiento es similar al efecto de los segundos momentos de la matriz Hessiana, Ecuación 1. Todo este proceso se detallará en la sección 3.2.

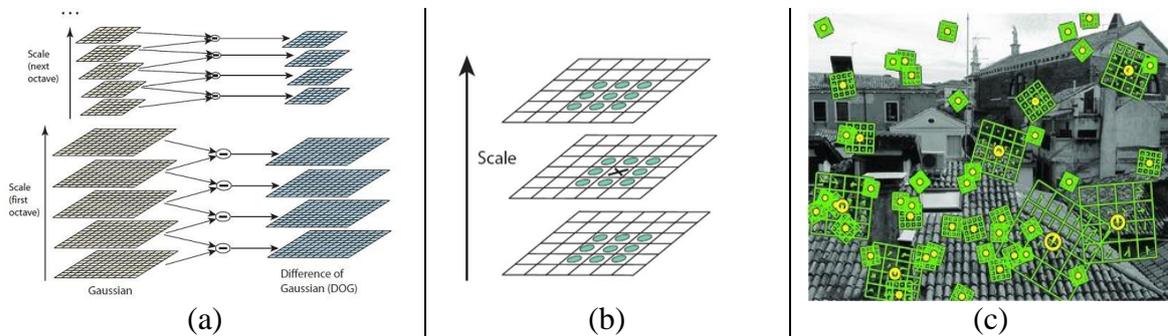


Figura 2-3: (a) Pirámides Gaussianas espacio escala y DOG, (b) análisis localizado y (c) resultado de detección. Fuente: a)[24], b)[24] y c) Propia

Dadas las notables propiedades de SIFT, las aplicaciones de su método tienen cabida en infinidad de ámbitos del procesamiento de imagen. Desde su proposición inicial en el paper de Lowe [24] para reconocimiento de objetos, su uso se ha ido extendiendo. Entre las aplicaciones más destacadas se encuentran:

- Reconocimiento de objetos con y sin oclusión[1]
- Categorización de objetos [10][20][21].
- Modelado, reconocimiento y seguimiento en vídeo[4]
- Generación de panorámicas[3].

En definitiva, los puntos SIFT han sido utilizados en muchos de los ámbitos del análisis de imágenes y video con mejores o peores resultados, efecto que se comprobará en la sección 3.2.

2.2.1.4 Detector de SURF

Al igual que el detector Hessiano resultaba similar al de Harris, los puntos SURF [19] presentan altas similitudes con los puntos SIFT. Usando una distribución de matrices Hessianas directamente, y reduciendo los procesos de SIFT a los elementos básicos y esenciales, se obtiene un método que mejora ampliamente el tiempo de procesado de su predecesor con similares resultados.

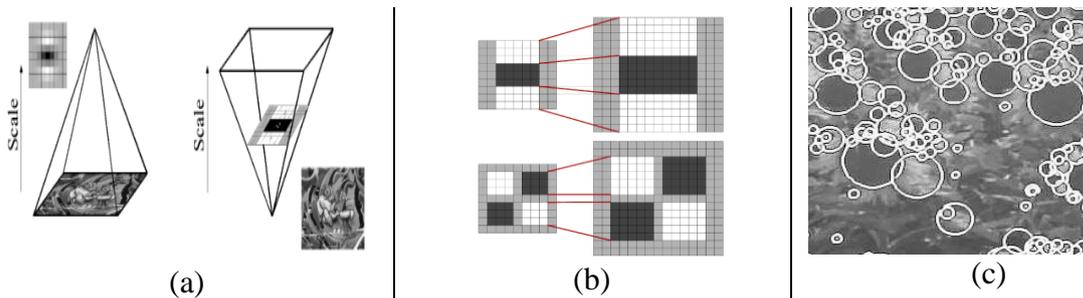


Figura 2-4: (a) Análisis multiescala, (b) normalización de espacio y (c) resultado de detección. *Fuente:* [19]

Su aplicaciones, pese a presentarse también en los mismos campos que SIFT, se centran más en la técnicas relacionadas con el vídeo, dada su gran ventaja en tiempo de procesado [25][26].

2.2.1.5 Detector de MSER

El algoritmo de detección de regiones MSER (*Maximally Stable Extremal Regions*) [17] sigue un proceso similar a un algoritmo de extracción de componentes conexas. La aplicación de dicho algoritmo junto con una serie de umbralizaciones de estabilidad desde los mínimos de la imagen hasta los máximos y viceversa permite obtener las regiones que presentan propiedades de invariancia afín, detección multiescala y estabilidad.



Figura 2-5: MSER detectadas con distintos umbrales de estabilidad. *Fuente:* [28]

La detección de estas MSER, propuesta por Matas [17], busca la detección de regiones con una serie de propiedades de robustez e invariancia que permitan la aplicación de técnicas de reconocimiento y seguimiento de objetos [28] o de búsqueda sobre grandes bases de datos de manera más eficiente [17].

2.2.2 Algoritmos de descripción

2.2.2.1 Descriptor de SIFT

Tal y como se comentó en la sección 2.2.2.1, los puntos SIFT [24] presentan numerosas propiedades. Algunas de ellas eran obtenidas gracias a la etapa de detección. Sin embargo, la descripción también resulta una etapa fundamental para el método. El proceso llevado a cabo consiste en el cálculo de la orientación principal del punto en la región detectada, la generación de un histograma de gradientes orientado según dicha orientación principal y la generación de un vector de descripción normalizado a partir de dicho histograma de gradientes.

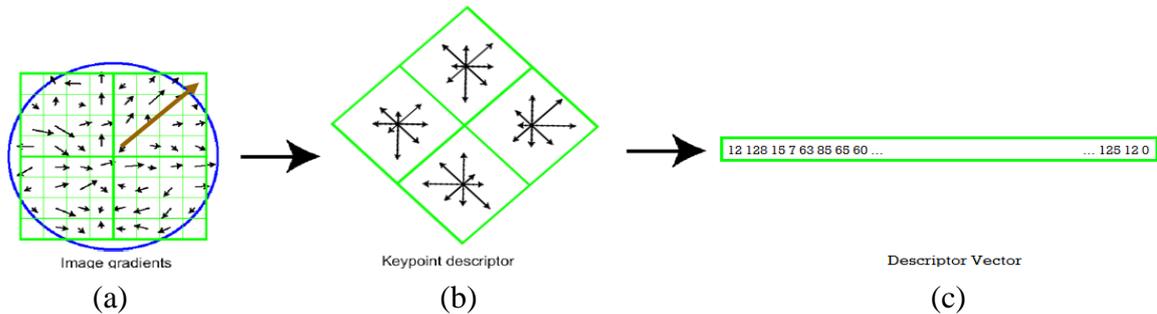


Figura 2-6: (a) Orientación principal, (b) gradientes orientados y (c) vector de descripción.

Fuente: a)[24], b)[24] y c) Propia

Las aplicaciones en las que SIFT es utilizado fueron ya expuestas en la sección 2.2.2.1.

2.2.2.2 GLOH

El descriptor GLOH (*gradient location and orientation histogram*) es, tal y como su nombre indica, un descriptor de localización de gradiente e histograma de orientación. Fue propuesto por Milkolajczyk y Schmid [14] como una extensión de SIFT con el objetivo de mejorar su robustez y distintividad. Se calcula el descriptor sobre un mallado log-polar dividido en secciones angulares y radiales que en ocasiones mejora el ajuste a la zona de interés del modelo cuadrado de SIFT.

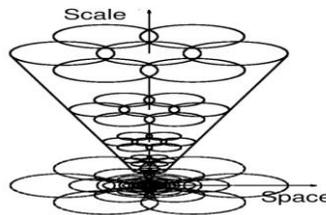


Figura 2-7: Modelo de análisis multiescala GLOH. Fuente: [46]

Las aplicaciones en las que se centra su utilización son por lo tanto las mismas sobre las que se aplican los puntos SIFT [2][14].

2.2.2.3 Descriptor basado en contexto de siluetas

El descriptor basado en siluetas [2] trabaja de manera similar a como lo hace SIFT, pero basándose en bordes. El denominado contexto de siluetas consiste en un histograma 3D de localizaciones de puntos de borde y sus orientaciones. Los puntos de borde se detectan mediante el detector de Canny [18]. Posteriormente, el descriptor es extraído de nuevo sobre una región log-polar.

Este método es principalmente utilizado en aplicaciones de reconocimiento de dígitos [2]. Sin embargo, ha sido, al menos evaluado, en alguna otra aplicación como son: recuperación de similitudes basadas en siluetas [29], reconocimiento 3D [30] o reconocimiento de logotipos [31].

2.2.2.4 Descriptor de imágenes de spin

La búsqueda de un descriptor basado en distribución, orientado al reconocimiento de objetos 3D, llevó a Johnson y Herbert a desarrollar este descriptor [32]. El descriptor almacena información sobre distancias al punto central e intensidad mediante interpolación bilineal. Los valores de intensidad se normalizan respecto del entorno (*patch*) y se cuantifican en 10 valores, desarrollando después un histograma normalizado de cada uno de los anillos entorno al punto central.

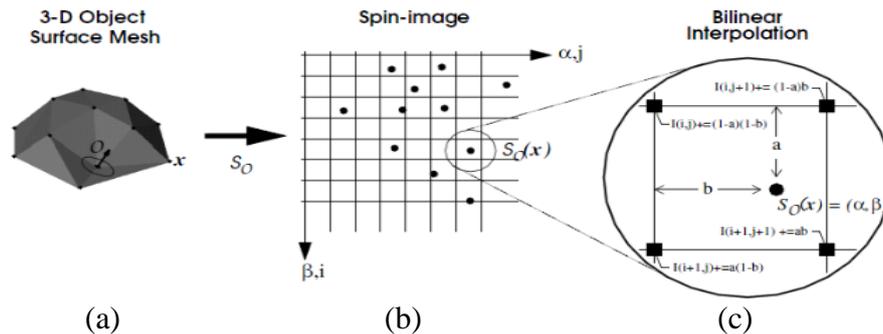


Figura 2-8: (a) Figura con detecciones, (b) puntos sobre la *spin image* y (c) modo de interpolación de la intensidad. Fuente: [32]

Aunque la aplicación objetivo de este descriptor está muy marcada por su origen, en los últimos años se ha conseguido adaptar su utilización para imágenes 2D [39] y su uso en estas como reconocedor de objetos y texturas [14].

2.2.2.5 Descriptor de momentos invariantes

Pese a desarrollar técnicas un poco alejadas de las tendencias comunes en el estado del arte, los resultados que ha ofrecido este descriptor han hecho que se haya incluido en el grupo de estudio. Van Gool propone un descriptor [33] para extraer la información de la naturaleza multispectral de la imagen. Extrayendo los momentos centrales definidos como:

$$M_{pq}^a = \iint_{\Omega} x^p y^q [I(x, y)]^a dx dy \quad \text{Ecuación 2}$$

de orden $p + q$ y de grado a , caracteriza las formas y las intensidades de una región Ω . Los momentos calculados en las direcciones cartesianas OX y OY serán los que compongan el descriptor.

Las aplicaciones objetivo de este descriptor son fundamentalmente el reconocimiento de objetos en diversas condiciones de iluminación y cambios del punto de vista [33][34], si bien es cierto que su funcionamiento queda muy limitado si se trabaja con imágenes en escala de grises, donde los momentos son menos discriminativos.

2.3 Comparativa de métodos por propiedades

Expuestos los métodos, se va a proceder a realizar una breve comparativa, que permita seleccionar qué algoritmo va a ser el que se pretenda mejorar en este proyecto. Para ello, se atenderá a los siguientes criterios:

- Resultados obtenidos en las evaluaciones más referenciadas del estado del arte[13][14][16].
- Grado de madurez del método y número de referencias y aplicaciones en las que ha sido desarrollado.
- Facilidad en implementación y modificación

Para la primera premisa, se recurrirá a los resultados obtenidos por los distintos detectores y descriptores en las citadas evaluaciones. En dicho artículos, se realizan una serie de comparativas por propiedades y resultados, que permitirán seleccionar el o los métodos más punteros del estado del arte.

2.3.1 Comparativa de detectores

Para la evaluación del comportamiento de los distintos detectores, Mikolajczyk [16] analiza todos los resultados obtenidos de los distintos detectores usando siempre el mismo descriptor (SIFT).

En el citado artículo se hace un estudio sobre tres propiedades:

- Invariancia a la rotación.
- Invariancia de escala.
- Invariancia afín.

Adicionalmente, se estudia el comportamiento de los detectores sobre los cuatro parámetros más empleados en la literatura para definir la calidad de un detector:

- Repetibilidad.
- Precisión de la localización.
- Robustez.
- Eficiencia.

Los resultados obtenidos han sido:

Detector	Esquina	Blob	Región	Inv. Rotación	Inv. Escala	Inv. Afín	Repet.	Precisión Localización	Robustez	Eficiencia
Harris – Affine	✓	✓		✓	✓	✓	+++	+++	++	++
Hessian – Affine	✓	✓		✓	✓	✓	+++	+++	+++	++
MSER			✓	✓	✓	✓	+++	+++	++	+++
DOG	✓	✓		✓	✓		++	++	++	++
SURF	✓	✓		✓	✓		++	++	++	+++

Tabla 2-1: Comparativa de detectores Fuente: [16]

Atendiendo a los resultados obtenidos por [16] plasmados en la tabla anterior, se pueden extraer las siguientes conclusiones.

- A nivel de propiedades: Los detectores de DOG y SURF no presentan la invariancia afín.
- A nivel de resultados: El detector de Hessian-Affine y el de MSER presentan las mejores puntuaciones.

Por último, cabe destacar que, de atenderse al tipo de elemento detectado, a nivel de esquinas, el detector de Harris-Affine quedaría como el mejor, a nivel de blobs el citado Hessian-Affine y a nivel de región las MSER.

Si bien es cierto que no se ha presentado el detector de SIFT como un elemento más de la tabla comparativa, el autor citado no lo incluye como tal atendiendo a dos motivos:

- El detector de SIFT no se desarrolló como tal en la bibliografía, sino como parte de un método compacto que incluye detección y descripción.
- Sus similitudes a nivel de proceso hace que sus resultados queden representados por el detector de Hessian-Affine en la gran mayoría de situaciones, como se expone en [15].

Así pues, se puede concluir que, como método genérico de detección de puntos, blobs o regiones características, el detector de Hessian-Affine (~SIFT) es el que presenta mejores prestaciones.

2.3.2 Comparativa de descriptores

Al igual que se ha procedido con los detectores, para la comparativa de los distintos descriptores se recurrirá a la realizada por el autor más referenciado en este ámbito[14].

En el citado artículo, se procede a comparar los distintos métodos de descripción sobre las detecciones realizadas por el mejor detector del estado del arte (Hessian-Affine).

Para llevar a cabo una evaluación exhaustiva, se procede a estudiar la Precisión y el *Recall* de los descriptores sobre imágenes que presentan las siguientes propiedades:

- Cambios de escala.
- Cambios de iluminación.
- Borrosidad.
- Rotación.
- Cambios de punto de vista.

El ranking obtenido mediante observación de los datos aportados en el artículo es el siguiente:

Detector	Cambio Escala	Rotación	Cambio Iluminación	Cambio Punto de Vista	Borrosidad
SIFT	2	1	2	2	2
GLOH	1	1	1	1	1
Contexto de Siluetas	3	2	3	3	3
Imágenes de Spin	4	5	5	4	4
Momentos Invariantes	5	3	4	5	5

Tabla 2-2: Comparativa de descriptores Fuente: [14]

Esta tabla ha sido elaborada a partir de los resultados observados en el punto de trabajo óptimo de las curvas de Precisión y *Recall* expuestas en el artículo.

A la luz de los resultados extraídos de la comparativa llevada a cabo por Mikolajczyk, y plasmada en la Tabla 2-2, se puede concluir que el descriptor de SIFT, o su modificación introducida en GLOH, presenta los mejores resultados frente a las distintas eventualidades estudiadas.

Se considerará por tanto dicho descriptor como el mejor del estado de arte de la descripción de imágenes mediante puntos, blobs y regiones características.

2.4 Conclusión del Estado del Arte

Tras el estudio llevado a cabo en este punto, y junto con los datos obtenidos de la bibliografía, se ha podido llegar a dos conclusiones a nivel de detectores y descriptores punteros en el estado del arte de la PROIID.

En el campo de la detección, por sus excelentes resultados en todas las propiedades estudiadas, así como por la alta diversidad de aplicaciones en las que ha sido empleado 2.2.1.2, el detector de Hessian-Affine, o el similar empleado en SIFT [24], es considerado el mejor detector del estado del arte actual, y será por tanto el detector objetivo de mejora del algoritmo a desarrollar¹.

Por otra parte, en el campo de la descripción, atendiendo únicamente a los resultados obtenidos en los artículos comparativos citados, el descriptor de GLOH presentaría leves ventajas frente a su predecesor SIFT.

Sin embargo, puesto que es objetivo de este proyecto evaluar el algoritmo a desarrollar frente a los métodos del estado del arte en distintas aplicaciones y propiedades, se tendrá en cuenta adicionalmente el segundo aspecto citado en el punto de comparativa de propiedades 2.3, el cual hace alusión a:

- Número de referencias e implementaciones del método que poder mejorar.
- Número de aplicaciones en las que ha sido utilizado sobre las que poder trabajar.

¹ En particular, por resultar más fácil trabajar con un método compacto sobre el que introducir modificaciones se trabajará con el detector de SIFT y no con el similar de Hessian-Affine. Éste mismo motivo vendrá a colación de la decisión de usar el detector de SIFT y no GLOH.

El reducido número de referencias y aplicaciones en las que el descriptor de GLOH ha sido citado, junto con las mínimas diferencias de comportamiento entre ambos, hará que se seleccione el descriptor de SIFT como objetivo de mejora del algoritmo propio.

En definitiva, será el **método de SIFT** (detector + descriptor) el objetivo de análisis, prueba y mejora de este proyecto.

Capítulo 3 – Análisis de detector y descriptor de SIFT

3.1 Introducción

El algoritmo *Scale Invariant Feature Transform* (SIFT) fue desarrollado por Lowe [1] como un método capaz de detectar y describir puntos característicos estables en una imagen. Estos puntos presentaban una serie de propiedades de estabilidad presentadas en el Estado del Arte Capítulo 2, que se estudiarán en detalle más adelante.

Estas propiedades han convertido al algoritmo en puntero en el estado del arte del procesado de imágenes mediante puntos o regiones de interés. Originalmente, este algoritmo fue desarrollado para el reconocimiento de objetos en imágenes estáticas [1], aunque sus aplicaciones han sido múltiples y numerosas desde entonces [4][5][12][20][21]. Para el estudio de sus características, capacidades y defectos, se va a seguir el esquema indicado en a continuación:

- En primer lugar, se analizará el funcionamiento del algoritmo.
- En segundo lugar, se extraerán las propiedades relevantes que surgen del proceso anteriormente descrito.

El algoritmo SIFT se compone principalmente de cuatro etapas que se describen siguiendo la implementación de Lowe:

1. **Detección de Extremos en el Espacio Escala:** La primera etapa del algoritmo realiza una búsqueda sobre las diferentes escalas y dimensiones de la imagen identificando posibles puntos de interés. Esto se lleva a cabo mediante un método basado en DOGs.
2. **Localización de los Puntos Clave:** Para seleccionar los puntos clave, también llamados puntos de interés, de forma precisa, se aplica una medida de estabilidad sobre todos ellos para descartar aquellos que no sean adecuados.
3. **Asignación de la Orientación Principal:** Se asignan una o más orientaciones a cada punto de interés extraído de la imagen basándose en las direcciones locales presentes en la imagen gradiente. Todas las operaciones posteriores son realizadas sobre los datos transformados según la orientación, escala y localización dentro de la imagen asignados en esta etapa.
4. **Descriptor del Punto de Interés:** La última etapa hace referencia a la representación de los puntos clave como una medida de los gradientes locales de la imagen en las proximidades de dichos puntos clave y respecto de una determinada escala.

Por último, y con el fin de fortalecer la estabilidad de los puntos obtenidos en este método, Brown y Lowe [35] propusieron una función 3D para eliminar aquellos puntos de bajo contraste o ubicación más sensible al ruido. Con el tiempo, este método ha pasado a incorporarse al algoritmo de SIFT como una etapa más pese a no encontrarse en el enunciado original.

3.2 Funcionamiento

En base a los puntos citados en la breve descripción llevada a cabo en la sección 3.1, se va a proceder a desarrollar más en profundidad cada uno de ellos, habida cuenta de que será éste el método base respecto al que se desarrollará el nuevo algoritmo.

3.2.1 Detección de Extremos en el Espacio Escala

En esta primera etapa, se busca localizar candidatos a puntos SIFT. Entre otras características que habrán de presentar, estos puntos deberán de ser repetibles, esto es, detectables bajo diversas condiciones de tamaño y puntos de vista.

Para esto, se realiza un estudio de la imagen original tras un procesado espacio-escala Gaussiano. Así se definirá el espacio-escala de la imagen original $L(x, y, \sigma)$ como la convolución de funciones Gaussianas 2D $G(x, y, \sigma)$ de distinto factor de escala (σ) con la imagen $I(x, y)$.

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad \text{Ecuación 3}$$

con (x, y) como coordenadas espaciales. El algoritmo utiliza la función DOG que se forma a partir de la derivada escalar de la Gaussiana escalada espacialmente. Esta función DOG $D(x, y, \sigma)$ se obtiene mediante la sustracción de escalas posteriores en cada octava:

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \quad \text{Ecuación 4}$$

donde k es una constante multiplicativa del factor de escala. La función DOG es utilizada por varias razones. En primer lugar porque es una función eficiente en cuanto a coste computacional se refiere: Las imágenes suavizadas $L(x, y, \sigma)$ son calculadas para la descripción de características en el espacio-escala, y por lo tanto, D puede obtenerse como una simple resta. Además, Mikolajczyk [36] asegura que los máximos y mínimos del Laplaciano de la Gaussiana respecto de una escala normalizada producen las características de imagen más estables en comparación con otras funciones como el Gradiente, el Hessiano o el Harris Corner Detector, pudiéndose aproximar el Laplaciano de la Gaussiana de escala normalizada mediante la función DOG.

Al conjunto de las imágenes Gaussianas suavizadas junto con las imágenes DOG se le llama octava. El conjunto de las octavas es construido mediante el muestreo sucesivo de la imagen original. Cada una de las octavas es a su vez dividida en un número entero de subniveles o escalas s . Una vez se ha procesado una octava completa, la primera imagen de la siguiente octava se obtiene mediante el muestreo de la primera de las imágenes de la octava predecesora con un valor de σ del doble respecto a la actual. Esto se traduce en una gran eficiencia del algoritmo para un número de escalas pequeño. El proceso descrito puede verse representado en Figura 3-1. Es importante tener en cuenta que la imagen original es expandida en el inicio del proceso para crear más puntos de muestreo que en la imagen original, por lo que la imagen resulta duplicada en tamaño antes de construir el primer nivel de la pirámide.

Dado que el espacio-escala $L(x, y, \sigma)$ representa la misma información a diferentes niveles de escala, el modo particular del muestreo permite una reducción de la redundancia.

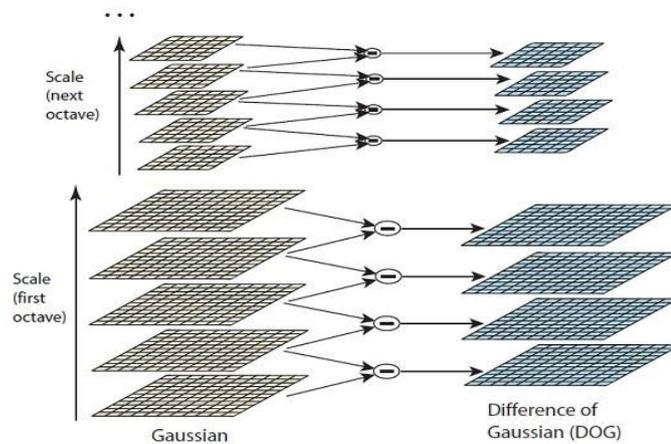


Figura 3-1: Espacio escala Gaussiano y DOG. Fuente: [24]

En cada una de las escalas, también llamadas octavas, la imagen se convolucionan repetidamente con funciones Gaussianas para producir el conjunto de imágenes Gaussianas mostradas en la parte izquierda de la figura. Las imágenes obtenidas son substraídas en parejas adyacentes para producir las imágenes DOG mostradas a la derecha. Después de cada octava, las imágenes Gaussianas son submuestreadas, y se repite el proceso.

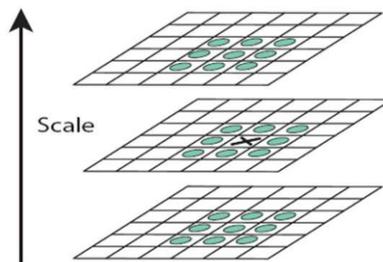


Figura 3-2: Localización de máximos y mínimos locales. Fuente: [24]

En la Figura 3-2 se ilustra como los máximos y mínimos de las imágenes diferencia-de-Gaussiana son detectados mediante la comparación de un pixel (marcado con X) con sus vecinos en las regiones de las escalas actual y adyacentes (marcados en azul). Si el valor resulta ser superior o inferior al de todos sus vecinos, se identifica el punto como máximo o mínimo local respectivamente.

3.2.2 Localización de puntos clave estables

Una vez los puntos clave candidatos han sido calculados, en esta segunda etapa se realiza un estudio de su estabilidad. Los puntos no firmemente situados sobre los bordes o aquellos con bajo contraste son bastante vulnerables al ruido y por lo tanto no podrán ser detectados bajo pequeños cambios de iluminación o variación del punto de vista de la imagen. Para excluirlos, Lowe propone los siguientes criterios:

- Para eliminar los puntos con bajo contraste, se aplica un proceso de umbralización por el cual los puntos cuyo valor sea menor que un umbral serán excluidos de la siguiente etapa por no considerarse suficientemente estables.
- Los puntos situados sobre bordes de manera difusa conllevan un alto grado de inestabilidad incluso ante pequeños ruidos. Para llevar a cabo su eliminación se utiliza la propiedad de la función DOG que presenta una gran curvatura en la dirección paralela al borde y pequeña en la dirección perpendicular cuando se encuentra en bordes difusos. Estas respuestas tan características se pueden estudiar mediante el cálculo de la matriz:

$$\mathbf{H} = \begin{bmatrix} \frac{\delta^2 D}{\delta x^2} & \frac{\delta^2 D}{\delta x \delta y} \\ \frac{\delta^2 D}{\delta x \delta y} & \frac{\delta^2 D}{\delta y^2} \end{bmatrix} \quad \text{Ecuación 5}$$

donde D es la imagen DOG $D(x, y, \sigma)$ respecto a la escala s . Las derivadas se calculan como la resta del valor de los puntos vecinos. Se puede demostrar que la siguiente desigualdad permite la localización de los puntos en los bordes:

$$\frac{\left(\frac{\delta^2 D}{\delta x^2} \times \frac{\delta^2 D}{\delta y^2}\right)^2}{\left(\frac{\delta^2 D}{\delta x^2} \times \frac{\delta^2 D}{\delta y^2}\right) - \left(\frac{\delta^2 D}{\delta x \delta y}\right)^2} = \frac{(r+1)^2}{r} \quad \text{Ecuación 6}$$

Así pues, aquellos puntos que no satisfagan esta desigualdad serán descartados debido a su inestabilidad. Tras descartar los puntos inestables, a todos los puntos se les dará una orientación.

3.2.3 Asignación de la orientación

La característica principal de los puntos SIFT es que éstos son invariantes a una serie de transformaciones sobre las imágenes. La invariancia respecto de la rotación se consigue mediante la asignación a cada uno de los puntos de una orientación basada en las propiedades locales de la imagen y representando el descriptor respecto de esta orientación. Para cada uno de los puntos de interés se calcula la magnitud del gradiente, m , y su orientación, θ , mediante las siguientes ecuaciones:

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \quad \text{Ecuación 7}$$

$$\theta(x, y) = \arctan \frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)} \quad \text{Ecuación 8}$$

donde L representa la imagen Gaussiana suavizada cuya escala resulta más próxima a la escala del punto de interés actual.

Respecto de la orientación del gradiente, se crea un histograma con un número determinado de *bins*, cada uno de ellos con una longitud determinada (20° en [24]) para cubrir el rango de los 360° posibles. El *bin* cuyo valor más alto se corresponde con la dirección dominante del gradiente y, por lo tanto es elegido como orientación principal.

Sin embargo se ha de tener en cuenta la posibilidad de que exista más de una dirección dominante. Es por ello que cualquier bin con un valor de más de 0.8 el valor de la magnitud principal (Lowe [24]) se considerará también como dirección dominante. Los puntos que contengan más de una dirección dominante supondrán una mayor estabilidad al mismo. Para una mayor precisión se utiliza una parábola para mediante la interpolación de varios de los valores más altos del histograma obtener el valor del pico.

Las orientaciones principales del histograma se asignan al punto de interés para que así el descriptor quede representado respecto de éstas.

3.2.4 Descripción del punto de interés

Las etapas anteriores han dotado a los puntos de interés seleccionados de invariancia respecto de la orientación, escala y localización. En esta última etapa se crea un vector de características para cada uno de los puntos de interés que contiene una estadística local de las orientaciones del gradiente de la escala de espacio Gaussiano. Se realiza un muestreo de las orientaciones y magnitudes del gradiente de la imagen sobre regiones alrededor del punto de interés. Este proceso es similar al de la etapa anterior, donde ahora cada una de las muestras son ponderadas tanto por la magnitud de su gradiente como por una función 3D Gaussiana evitando así cambios bruscos en el descriptor ante pequeños cambios en la posición de la ventana y al mismo tiempo asignando menor énfasis a los puntos más alejados del punto de interés.

Se analizan las muestras de cada región formando histogramas de orientaciones resumiendo el contenido en sub-regiones como se puede ver en la Figura 3-3: Descripción de los puntos de interés. Cada uno de los histogramas se compone de N bins (N=8 propuesto por Lowe [24]), que almacenan las orientaciones posibles proporcionales. La magnitud de cada flecha representa el valor acumulado para cada bin. Por lo tanto se obtienen un determinado número histogramas (4x4 Lowe [24]) respecto de las orientaciones de los puntos de cada región para cada uno de los puntos de interés.

Finalmente el descriptor de cada punto de interés está formado por un vector que contiene los valores de todas las orientaciones de cada uno de los histogramas, componiendo un vector de características.

Adicionalmente, el vector de características es normalizado para dotarlo de cierta robustez frente a cambios de iluminación. Los cambios de iluminación afectan en mayor medida a la magnitud del gradiente y no a la orientación, por lo que se busca una representación de esta magnitud que minimice estos efectos.

Para ello se lleva a cabo un proceso de normalización en los que ahora los cambios de contraste (píxeles multiplicados por una constante) quedan neutralizados, mientras que los cambios en la luminosidad (suma de una constante con los píxeles) no afecta a los valores del gradiente que se calcula como diferencias entre píxeles.

Si bien esta normalización no confiere invariancia respecto de los cambios de iluminación, si se consigue paliar parcialmente los efectos que estos producen.

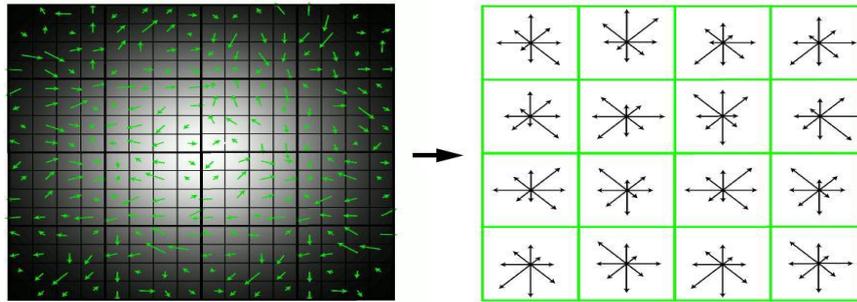


Figura 3-3: Descripción de los puntos de interés Fuente: [24]

Finalmente se limita el valor de cada componente de magnitud de gradiente a un valor máximo para que tenga un mayor peso la orientación frente a la magnitud del gradiente. Luego se vuelve a normalizar de nuevo a una amplitud unidad.

3.3 Propiedades

Como resultado de cada uno de los pasos seguidos en el algoritmo descrito, se obtienen unos puntos de interés cuya detección y descripción presenta una serie de propiedades que se explicarán y ejemplificará² a continuación.

3.3.1 Invariancia a cambios de escala

Durante la explicación del funcionamiento del algoritmo, se explicó cómo los puntos eran detectados sobre imágenes a las que se iban aplicando cambios de tamaño (octavas), lo cual permite que se detecten puntos a distintos tamaños de un mismo objeto.

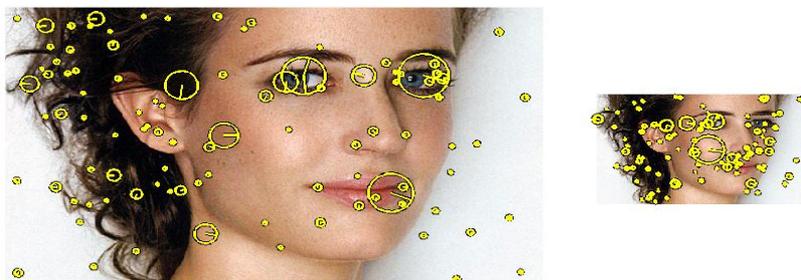


Figura 3-4: SIFT sobre cambio de escala. Fuente: Propia

Este proceso permite que los puntos detectados resulten ser invariantes a los cambios de tamaño, debido a que un punto detectado en varias octavas lo será también si el objeto inicialmente presenta un tamaño menor, pues esto equivale, salvo efectos del suavizado y submuestreo, a buscarlo en una octava menor.

3.3.2 Invariancia a rotaciones

Para cada punto detectado se obtenía la orientación principal en función del gradiente en su entorno. A partir de ese punto, todos los cálculos restantes se llevaban a cabo en función de dicha orientación extraída.

² Imagen de ejemplo extraída de: Eva Green (cropped scan from Tatler December 2006)

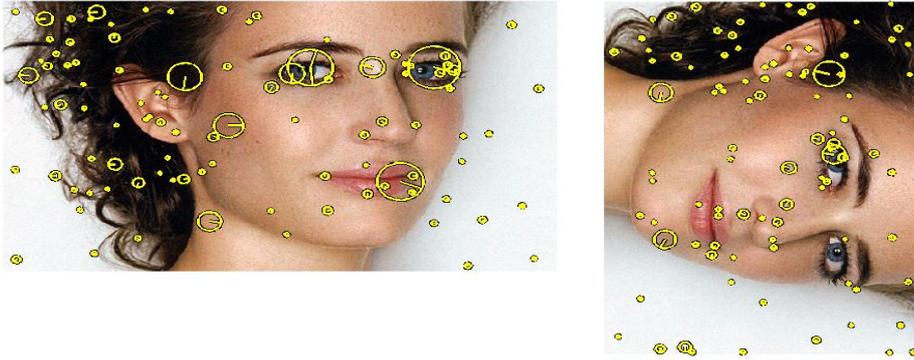


Figura 3-5: SIFT sobre rotación. Fuente: Propia

Por lo tanto, si el entorno de un punto se mantiene estable, esto es, rota con él, la orientación relativa del punto seguirá siendo la misma, y por tanto los cálculos realizados a partir de ahí dará lugar a idénticos resultados, habiéndose obtenido pues la citada invariancia a rotaciones.

3.3.3 Invariancia a cambios lineales de iluminación

Una vez se ha extraído el descriptor, la normalización del mismo permite que, de tener lugar cambios lineales de luminosidad, estos no afecten al resultado final y por tanto se obtengan descriptores de puntos con invariancia a los cambios lineales de luminosidad.

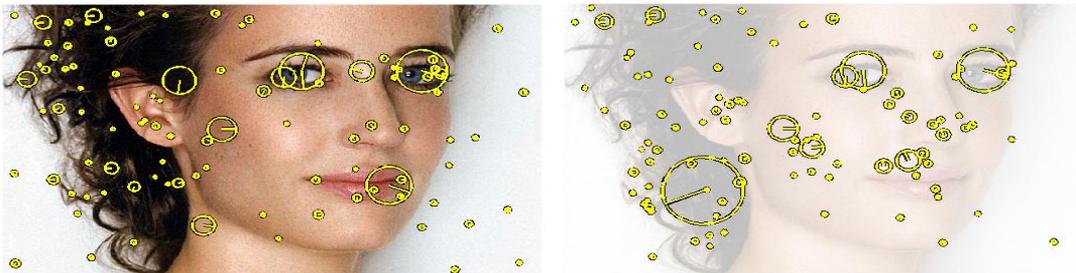


Figura 3-6: SIFT sobre iluminación lineal. Fuente: Propia

3.3.4 Invariancia a borrosidad

Dentro de cada octava en la que se buscaban los puntos SIFT, se realizaba también una división en escalas, esto es, se buscaban los puntos sobre imágenes del mismo tamaño (octava) pero en distintos niveles de suavizado (escalas).



Figura 3-7: SIFT sobre borrosidad. Fuente: Propia

Así, los puntos que eran detectados bajo estas condiciones resultaban ser invariantes a la borrosidad de las imágenes.

3.3.5 Invariancia al ruido

Tal y como se comentó en el apartado de funcionamiento, Brown y Lowe incluyeron una función 3D con el fin de eliminar puntos de bajo contraste, o detectados en bordes de baja estabilidad. Estos puntos resultaban ser muy sensibles al ruido, y la función de Brown y Lowe permitía eliminarlos, dejando únicamente como resultado aquellos que sí eran invariantes frente al ruido³.

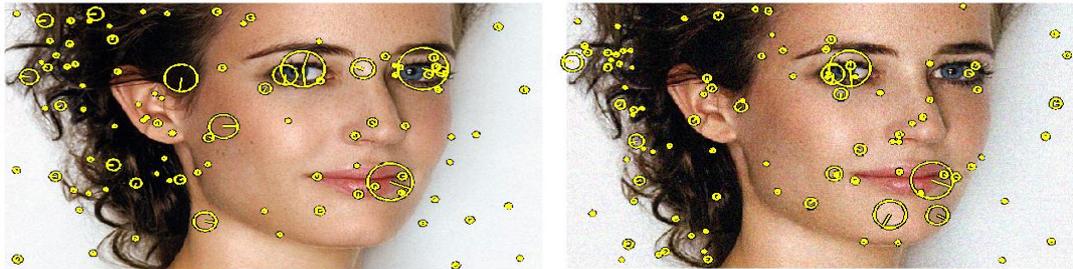


Figura 3-8: SIFT sobre ruido. Fuente: Propia

3.3.6 Invariancia a transformaciones afines

El método de descripción de SIFT, esto es, la extracción de histogramas sobre regiones del entorno divididas en cuadrantes y a su vez en secciones de 4x4 permite la generación de una información de punto invariante a las posibles transformaciones afines sufridas por el objeto. Las transformaciones afines resultan ser cambios afines del punto de vista.



Figura 3-9: SIFT sobre transformación afín. Fuente: Propia

3.3.7 Alta distintividad

Por último, estos 128 valores obtenidos para cada punto, presentan una muy elevada distintividad gracias al proceso seguido de detección y descripción en función del entorno y sus propiedades poco repetibles.

³ Ruido de tipo *salt&pepper* de densidad $d=0.25$

Capítulo 4 – SIFT: Defectos de Aplicación

4.1 Introducción

Habida cuenta de las citadas propiedades presentes en los puntos SIFT 3.3, las aplicaciones para las que han sido utilizados han sido muchas y muy variadas. A continuación se va a incluir un estudio de las aplicaciones más relevantes en las que se han usado y como de positiva o negativamente afectan las antes mencionadas propiedades de SIFT a su comportamiento en las mismas

Una visión general de las aplicaciones en las que se han empleado SIFT, y de las propiedades que requieren es la siguiente:

Matchings sobre grandes bases de datos	• Detectores repetitivos • Descriptores "ligeros" y distintivos	[38]
Reconocimiento de objetos y <i>duplicate images</i>	• Detectores repetitivos en posición y contenido • Descriptores distintivos y que independicen objetos	[2][37]
Reconocimiento de texturas.	• Detectores no muy numerosos • Descriptores muy distintivos	[39]
<i>Reconocimiento de objetos ocluidos</i>	• Detectores repetitivos en contenido • Descriptores que independicen objetos	[40]
Construcción de panorámicas.	• Detectores de regiones estables • Descriptores con invarianza afin	[3]
Minería de datos en vídeo.	• Detectores repetitivos • Descriptores con invarianza afin y distintivos	[4]

Figura 4-1: Esquema de aplicaciones con SIFT y propiedades requeridas

Tal y como se observa en la figura anterior, los puntos SIFT parecen satisfacer las principales necesidades de las aplicaciones, salvo en dos casos:

- Reconocimiento de objetos y *duplicate image*.
- Reconocimiento de objetos ante oclusiones

En ambos casos, una de los principales requisitos exigidos es que el método sea capaz de aislar la información correspondiente al objeto, de aquí en adelante denominado como frente, de la del fondo, mientras que SIFT trata de describir cada punto con toda la información del entorno.

Habida cuenta de que en estas dos aplicaciones se está dando un uso a SIFT que no corresponde con sus propiedades, se van a analizar en detalle con la intención de localizar puntos en los que sea posible mejorar su comportamiento, esto es, ese va a llevar a cabo una búsqueda de defectos de aplicación.

4.2 Reconocimiento de objetos e imágenes duplicadas

El reconocimiento de objetos o de *Duplicate Image*, como logotipos o iconos,[2][37] consiste en la extracción de información de un objeto presente en una escena, y la comparación de la misma con la extraída de otra escena diferente en la que se evalúa si se encuentra de nuevo dicho objeto.

Si bien es cierto que ya en el artículo original donde se propuso SIFT [1] se incluye la aplicación de reconocimiento de objetos como aplicación demostrativa, se trata de un reconocimiento bajo unas condiciones de entorno muy controladas.

Sin embargo, el reconocimiento de objetos en ámbito genérico debería de permitir una serie de situaciones que se van a exponer a continuación:

- Funcionamiento correcto bajo cambios controlados sobre el propio objeto:
Los puntos SIFT presentan una serie de propiedades ya expuestas con anterioridad, que los convierten en idóneos para este fin. Sus capacidades de invariancia frente a todo tipo de cambios permiten que, siempre que los cambios sobre el objeto en cuestión sean controlados, las descripciones sobre el mismo resulten muy similares.
- Funcionamiento correcto frente a cambios controlados o no del entorno:
Los puntos SIFT presentan un grave defecto en esta funcionalidad. Tal y como se expuso en la sección 3.3, una de las principales fortalezas de los puntos SIFT radica en su capacidad de describir una característica de la imagen mediante la información de su entorno. Esta propiedad resulta útil en la gran mayoría de casos, mas no en este.

El hecho de incluir en la descripción de un objeto información del entorno puede dar lugar a dos situaciones con consecuencias dispares:

- Si la detección tiene lugar sobre el objeto, la información de entorno resulta extremadamente útil, por lo que resulta muy positiva esta propiedad.
- Si la detección tiene lugar en el borde del objeto, se incluirá información del fondo en la descripción del objeto, lo cual resultará muy negativo, principalmente en casos como:
 - Entornos muy cambiantes o con mucha información (muy texturados).
 - Objetos con poca información caracterizados principalmente por información del borde-fondo.

Pese a que, como podría parecer, una situación contrarresta a la otra, es útil recordar de nuevo, que la mayoría de detecciones de SIFT tienen lugar en los bordes, por lo que la segunda situación resulta mucho más relevante.

Esta situación resulta por lo tanto un grave problema para el reconocimiento de objetos mediante puntos SIFT. Aun así, es cierto que los puntos SIFT siguen mejorando el comportamiento de la gran mayoría de métodos.

Sin embargo, si el reconocimiento de objetos se lleva a cabo de un modo genérico, dándose situaciones de oclusión parcial de los objetos o fondos muy texturados, este defecto se potencia aún más, pues la información del entorno influye de manera más notable en la descripción del objeto.

Así pues, y a modo de conclusión de este apartado, se puede exponer que SIFT presenta defectos en la descripción de objetos cuando incluye información del entorno, especialmente en situaciones de oclusión y entornos cambiantes o con mucha información. Esto imposibilitará su correcto funcionamiento en aplicaciones de reconocimiento de objetos y de categorías de objetos.

4.3 Reconocimiento de objetos ante oclusiones

Las oclusiones se definen en tratamiento de imágenes como la interposición entre el visor y el frente de elementos de la escena que impiden su visualización parcialmente. Las aplicaciones que tienen como objetivo el reconocimiento de objetos [3] o su seguimiento [40], encuentran un gran reto cuando estas situaciones tienen lugar.

Estas aplicaciones, como tantas otras, han sido un campo en el que los puntos SIFT han sido utilizados. Sin embargo, sus resultados no han sido del todo satisfactorios, y en ocasiones han sido necesarias introducir modificaciones en el algoritmo SIFT para su correcto funcionamiento en la aplicación concreta [40].

Como principales requerimientos de un método utilizado para seguimiento de objetos ante oclusiones destacan:

- Alta disitintividad en el reconocimiento genérico

Tal y como se comentaron en las propiedades de SIFT 3.3 y la figura de aplicaciones y propiedades Figura 4-1, las características de SIFT lo convierten en un método ideal para este tipo de aplicaciones.

La detección de puntos sobre zonas texturadas genera numerosas ubicaciones pero inferiores a las producidas por otros métodos expuestos en el estudio del estado del arte, lo cual permite una descripción computacionalmente más sencilla. Adicionalmente, los descriptores de SIFT resultan ser muy distintivos también en zonas de baja información, por lo que se cumplirá perfectamente con los requisitos de la aplicación en este aspecto.

- Capacidad de aislamiento de elementos independientes de una escena

Al igual que se mencionó en la sección anterior 4.2, los puntos SIFT tienden a incorporar a la descripción del punto o región de interés información del entorno. En este caso, el objetivo de la aplicación es la localización de objetos cuyas descripciones cambian por la interposición de elementos de la escena entre el visor y el objeto. Por lo tanto, la incorporación a la descripción de este segundo de información de elementos ocluyentes o fondos adyacentes resultará notablemente contraproducente.

Se extrae por lo tanto a nivel de esta aplicación, que las buenas propiedades del método de SIFT en cuanto al reconocimiento de objetos se ven mermadas por los defectos de descripción introducidos con las oclusiones. Este hecho imposibilitará por tanto su correcto funcionamiento en este tipo de aplicaciones. Cabe destacar, que los defectos asociados a esta aplicación no son tan notables como en el caso anterior.

Capítulo 5 – Algoritmo propuesto: RegSIFT

5.1 Introducción

Es cierto que la propiedad fundamental de SIFT de detectar y describir los puntos característicos de una imagen en función de un determinado vecindario le otorga una serie de cualidades muy destacadas como invariancia a los cambios de iluminación, invariancia a la rotación e incluso invariancia a las transformaciones lineales surgidas por cambios del punto de vista (sección 3.3).

Como se analizó en el Capítulo 4, esta propiedad, según la aplicación objetivo para la que sea usado el método puede llegar a convertirse en una debilidad. Para ilustrar este argumento se va a recurrir a un ejemplo de detección de objetos sobre distintos fondos (situación que puede darse en aplicaciones de SIFT tales como las presentadas en el Capítulo 4). El problema radica en que, para describir un objeto, por ejemplo, mediante los puntos SIFT que se detectan en su contorno, se incluye información en el descriptor del fondo sobre el que se encuentra. La Figura 5-1 presenta un ejemplo del comportamiento de SIFT en estos casos.

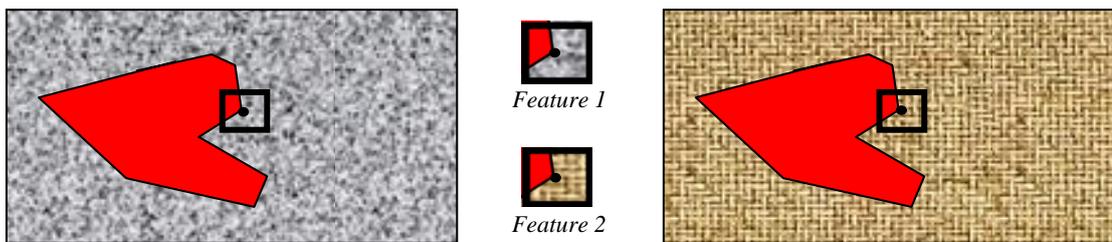


Figura 5-1: Detalle de dos *features* de SIFT

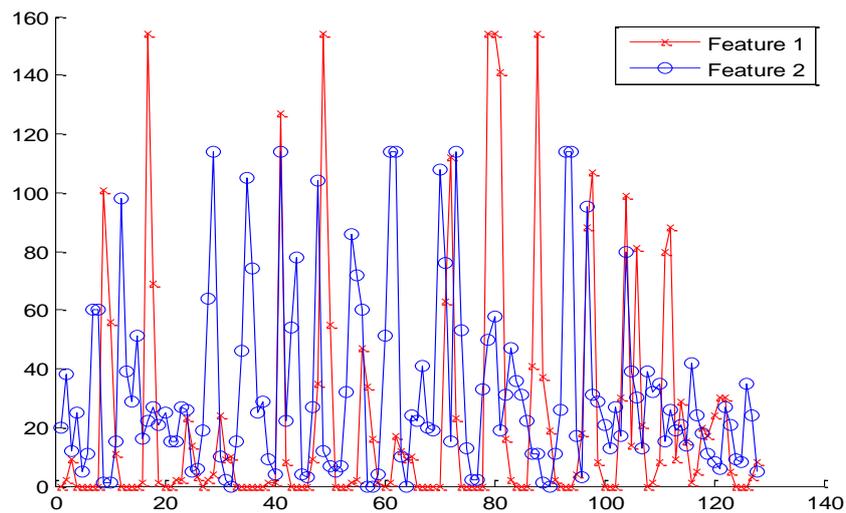


Figura 5-2: Comparación de descriptores de *Feature 1* y *Feature 2*

Se puede observar cómo, debido a la información añadida por el fondo, la misma posición del frente resulta dar dos descriptores completamente distintos.

Con el fin de solventar esta comportamiento indeseado (en función de la aplicación) de SIFT, se ha desarrollado el método denominado *RegSIFT*, que trata de aprovechar el hecho de que una segmentación en regiones tiende a respetar los bordes de los objetos en la escena. De esta manera, su inclusión en el proceso de SIFT permitirá corregir los defectos encontrados, sin hacerle perder sus características positivas, como se verá en la sección 6.3.

Para ello, en este método, se ha caracterizado cada punto SIFT no con el área total de su descriptor sino únicamente con la parte que pertenezca a una misma región detectada dentro del área del descriptor de SIFT. En definitiva, el método *RegSIFT* introduce las regiones en la etapa de descripción de los puntos SIFT con el fin de aislar la información del objeto a caracterizar de su entorno. Así, en el ejemplo de la figura anterior se obtendría un efecto como el mostrado a continuación.⁴

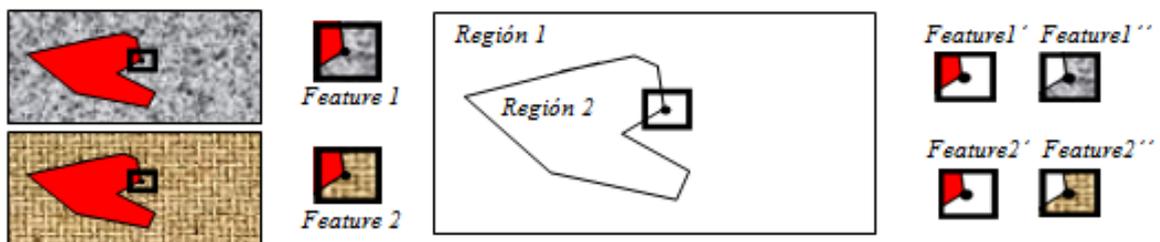


Figura 5-3: Ejemplo de *features* a describir en *RegSIFT*

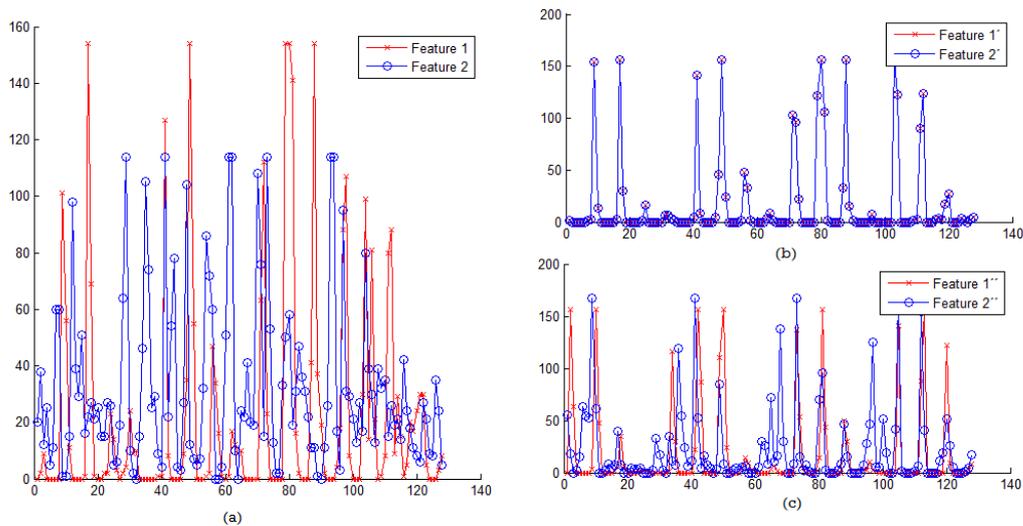


Figura 5-4: Descriptores de (a) *Features 1* y *2* de SIFT original, (b) *Feature1'* y *Feature2'* y (c) *Feature1''* y *Feature2''* de *RegSIFT*

En esta ocasión, se puede observar cómo, aislando la información de frente y fondo, la región de frente da lugar a descriptores idénticos y por tanto sí pueden ser machedados correctamente.

⁴ El ejemplo es ilustrativo. Nótese que la partición en regiones atendiendo al color resultaría en la creación de múltiples regiones en el fondo que empeorarían la claridad del ejemplo.

A continuación, se presentarán las modificaciones realizadas sobre las etapas de detección y descripción, sección 3.2. Dentro de ellas, se irán desglosando tanto las nuevas funcionalidades desarrolladas como las modificaciones introducidas en las ya existentes.

5.2 Algoritmo de detección

El proceso de detección seguirá dos caminos paralelos. Por una parte, el método de SIFT procederá como en el algoritmo original, a detectar los puntos característicos de la imagen, mientras que de manera simultánea⁵, se segmentará la imagen en regiones.

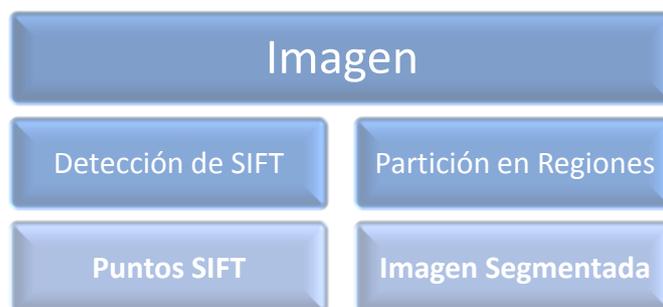


Figura 5-5: Esquema de detección de *RegSIFT*

Cada uno de los procesos involucrados en esta etapa de detección se comentará con un nivel superior de detalle en los epígrafes siguientes.

5.2.1 Detección SIFT

El proceso de detección de puntos característicos de SIFT se lleva a cabo de manera idéntica a como se describió en la sección 3.2 correspondiente a la detección. En él se trabaja sobre la imagen sin afectar al método original, y se obtienen los puntos tal y como lo haría el algoritmo original.

5.2.2 Partición en regiones

Tal y como se comentó en el Capítulo 4, los defectos de SIFT a la hora de trabajar con esas aplicaciones radican en que el área de descripción sobre el que trabaja no presenta ningún tipo de filtrado o división que le permita distinguir entre regiones a describir dentro de su área de descripción. Así, puntos en los bordes incluirán en la descripción información de ambas regiones delimitadas por dicho borde, creándose un error cuando la información de alguna de las regiones no es relevante.

Se buscará por tanto influir sobre SIFT de tal manera que se logre describir únicamente con el área del descriptor perteneciente a una misma región. Así, se conseguirán aislar las distintas informaciones pertenecientes a distintas regiones dentro de una misma área de descripción, logrando que la información irrelevante no camufle a la relevante.

Para lograr este cometido, se buscará llevar a cabo divisiones en la imagen con las siguientes propiedades características:

⁵ Idealmente, en la implementación realizada el proceso resulta ser secuencial.

- División de la imagen completa, sin exclusiones.
- Sin solape de regiones.
- Ajustadas a los bordes.

Si bien es cierto que en el apartado del estado del arte se analizaron distintos detectores de regiones como el detector de regiones de Harris, sección 2.2.1.1, o el detector de regiones MSER, sección 2.2.1.5, todas ellas tienden a seleccionar ciertas regiones y a descartar otras. En su mayoría presentan solapes, y por tratarse de regiones de estabilidad como en el caso de MSER, tienden a evitar los bordes por ser las zonas más inestables.

Habida cuenta de lo expuesto en estos párrafos, se habrá de recurrir a técnicas de segmentación, que resultar estar más acorde con las características requeridas. Antes de pasar a seleccionar un segmentador, cabe destacar las propiedades de los segmentadores de manera breve:

- Únicamente dividen la escena, sin llevar a cabo selecciones ni exclusiones.
- Al tratarse de una simple división no es posible el solape de regiones.
- En función de las propiedades más específicas de cada segmentador, se ajustarán en mayor o menor medida a los diferentes bordes, salvo en situaciones específicas como el camuflaje.

Con la técnica decidida, queda únicamente elegir el segmentador a utilizar. Para ello, se ha buscado en el estado del arte de los segmentadores uno que permita generar agrupaciones de píxeles lo suficientemente grandes para cobrar sentido semántico en comparación con un píxel, pero lo suficientemente reducida como para llevar a cabo un buen ajuste a los bordes.

Este tipo de segmentador basado en agrupaciones de píxeles de bajo nivel en regiones atómicas es referido por Mori [40] y por Ren and Malik [42] como *Superpíxeles* (SP).

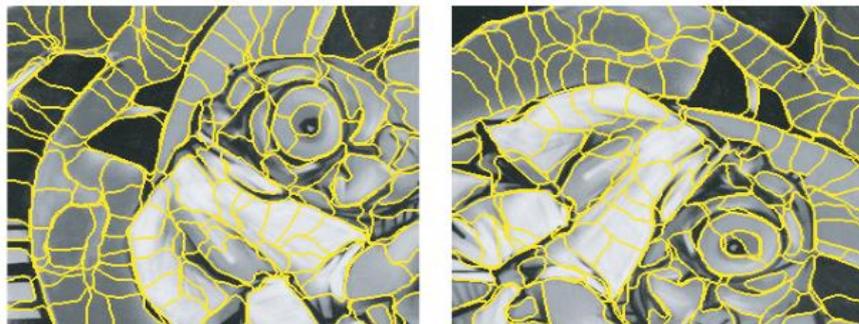


Figura 5-6: Ejemplo de segmentación por *Superpíxeles*

Sin embargo, y pese a que esta técnica cumple todos los requisitos expuestos, presenta una serie de defectos. Los dos más relevantes son expuestos por Tuytelaars y Mikolajczyk en [16] y se citarán brevemente.

- Baja distintividad:

El hecho de que se obtengan regiones homogéneas en toda la imagen hace que, por sí mismas, las regiones generadas sean poco distintivas y por lo tanto inútiles para aplicaciones como el reconocimiento de objetos.

Sin embargo, el hecho de que se vaya a utilizar como complemento para SIFT y no como método principal hace que este hecho se convierta en irrelevante.

- No invariancia a los cambios de escala:

Por la misma causa que el punto anterior. Todas las regiones extraídas en la imagen son homogéneas y además se obtienen directamente de dicha imagen original.

Puesto que el segmentador se va aplicar sobre la descripción de SIFT, la cual usa áreas de descripción extraídas a distintas escalas, se procederá a ir extrayendo los *Superpíxeles* para cada octava en la que trabaje SIFT. Se obtendrán por lo tanto pirámides de imágenes segmentadas.

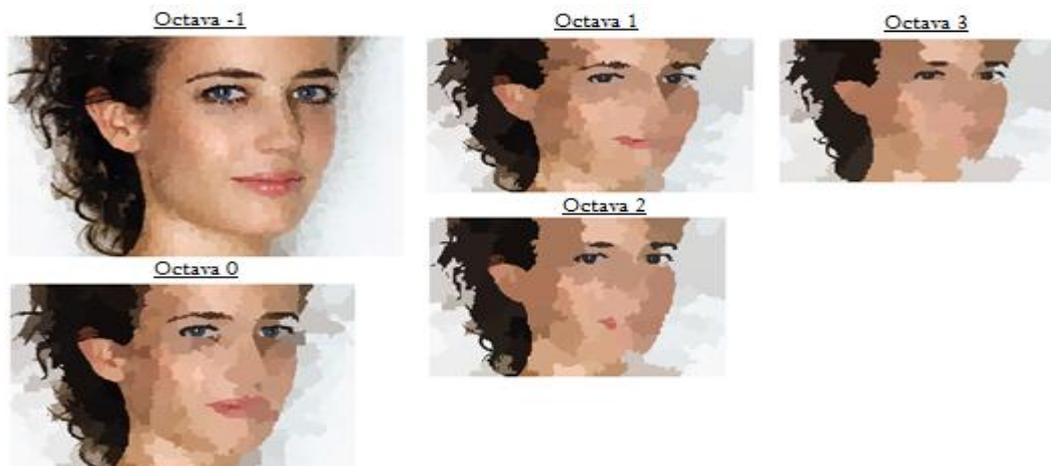


Figura 5-7: Ejemplo de *Superpíxeles* extraídos en diferentes octavas

En definitiva, como se ha podido ir viendo hasta este punto en la memoria, se puede decir que:

- Puntos SIFT: Pese a ser por si mismos el método más referenciado y con mejores propiedades del estado del arte, presenta notables defectos en ciertas aplicaciones, ver Capítulo 4.
- Segmentadores de *Superpíxeles*: Presentan unas grandes propiedades como segmentador, más no como descriptor de imágenes, como se ha expuesto en las líneas superiores.

Sin embargo, la combinación de SIFT + SP, permitirá obtener un algoritmo que presente únicamente los aspectos positivos de ambos, y por tanto un comportamiento superior a los obtenidos hasta la fecha en el estado del arte.

Una vez se han seleccionado los *Superpíxeles* como herramienta se pasa describir su actuación en el algoritmo.

De manera independiente a la detección de SIFT, se segmenta la imagen mediante los *Superpíxeles* en las distintas octavas que presente SIFT, Figura 5-7: Ejemplo de *Superpíxeles* extraídos en diferentes octavas. Para este ejemplo, y para el algoritmo desarrollado se ha utilizado la implementación de Vedaldi [43].

Esta etapa recibe pues la imagen de análisis y el número de octavas usadas por SIFT, y devuelve una pirámide de imágenes segmentadas.

5.3 Algoritmo de descripción

La descripción será, para este método implementado, la etapa en la que se influirá en el algoritmo original de SIFT. Mediante el uso de las regiones extraídas en la etapa de detección, se modificará la descripción que realiza SIFT sobre los vecindarios de los puntos de tal manera que se logren obtener *features* de las características deseadas para determinada aplicación (Capítulo 4).

El esquema de intervención se asemejará a lo siguiente:

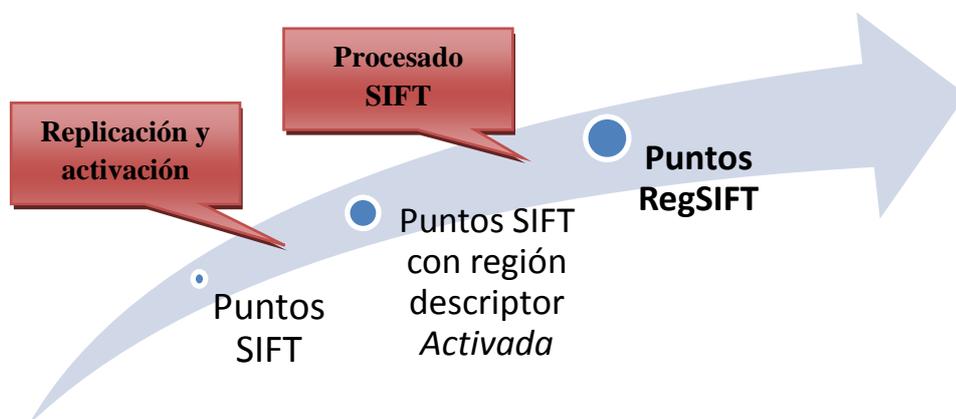


Figura 5-8: Esquema de descripción de RegSIFT

5.3.1 Replicación de *features*

La replicación de *features* será rediseñada respecto a la etapa 3.2.3 del algoritmo inicial de SIFT donde únicamente se replicaban los puntos con más de una orientación dominante. El objetivo de esta etapa es generar *features* idénticas en cuanto a su localización, con el fin de que cada una pueda describirse mediante las distintas regiones que hayan podido ser detectadas en el área de su descriptor.

Este bloque recibirá de la etapa de detección cada uno de los puntos característicos de SIFT detectados, los parámetros necesarios para calcular sus futuras áreas de descripción y una pirámide de imágenes resultado de la segmentación por escalas explicada en el punto 5.2.2 para la evaluación de los puntos “masqueados” por las regiones.

Con estos datos, esta etapa se llevará a cabo el chequeo de las regiones presentes en el área de descripción de cada punto. Dicho punto se replicará tantas veces como regiones de tamaño igual o superior al porcentaje umbral (OTh%) del área del descriptor se encuentren.

En función de estas regiones, se generan unas máscaras de descripción, *md*, asociadas a cada punto replicado. Estas máscaras se generarán como:

$$f_{mascara}(pixel)_i = \begin{cases} 1, & pixel \in R_i \\ 0, & resto \end{cases} \quad \forall i = 1 \dots P \quad \text{Ecuación 9}$$

con i barriendo las P regiones detectadas en el área del descriptor.

La figura siguiente muestra un detalle de la replicación desarrollada y las máscaras generadas:

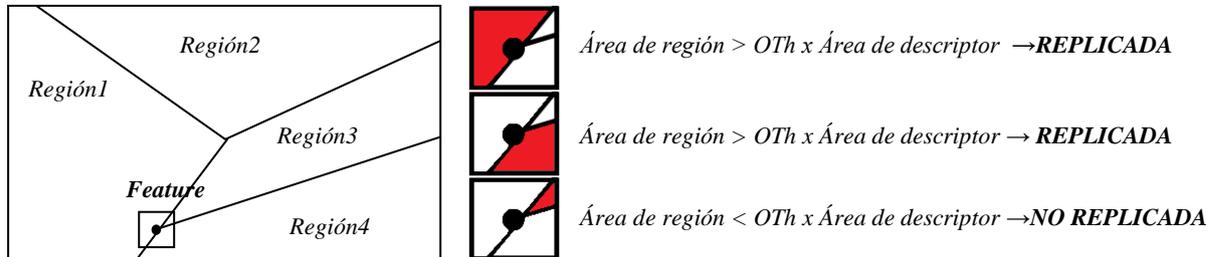


Figura 5-9: Ejemplo de replications por área de región y máscaras generadas

Cada *feature* de SIFT replicada, esto es, cada nueva *feature* de *RegSIFT*, presentará un nuevo área sobre el que extraer el descriptor, que vendrá definida por la zona *activada* (máscara) dentro del anterior área de descripción de SIFT.

Se entenderá por zona *activada*, o zona del *descriptor activada*, la parte del descriptor de SIFT original ocupado por una región, y que provocó la aparición de esta nueva *feature* (regiones en rojo)

Figura 5-9).

Como resultado final de esta etapa, se obtienen una serie de *features* con M nuevas regiones de descripción definidas por las distintas zonas *activadas* ($M \leq P$) dentro del descriptor original de SIFT.

5.3.2 Orientación principal

Con el fin de obtener invariancias a rotaciones y traslaciones lineales, tal y como se explicó en la sección de propiedades 3.3, los puntos SIFT calculan la orientación principal de cada punto para posteriormente generar el descriptor tomando como referencia dicha orientación. En esta ocasión, se influirá sobre dicha orientación, haciendo que únicamente se calcule en función de la nueva región de descripción.

Este bloque operativo contará con idénticas entradas que el bloque de replicación, salvo porque el número total de *features* corresponderá a las resultantes de la fase de dicha fase.

Para el cálculo de dicha orientación, se procederá a actuar de igual manera que lo hace SIFT (sección 3.2.3) pero anulando el gradiente en las zonas del área del descriptor de SIFT que no se correspondan con la del descriptor de *RegSIFT*.

Como resultado de esta etapa, se obtienen unas orientaciones principales para cada *feature* caracterizadas por la información contenida en sus regiones y el borde de gradiente que se forma. El efecto generado podría representarse una forma similar a lo mostrado en esta figura.

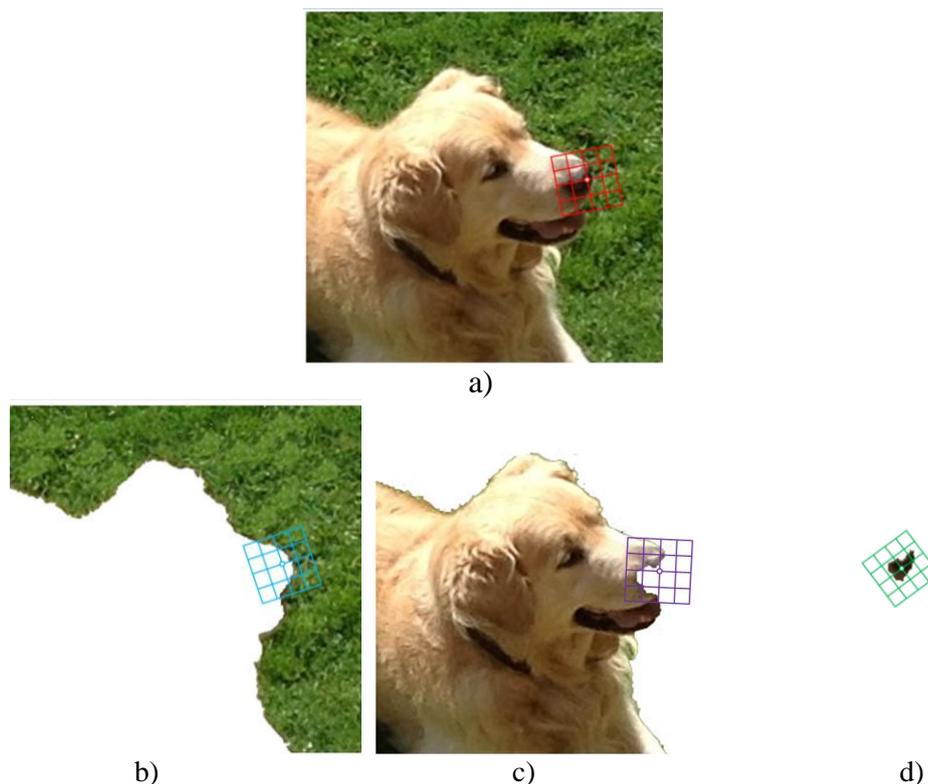


Figura 5-10: Detalle de orientación del descriptor. a)Descriptor SIFT b),c) y d) Descriptores RegSIFT. Fuente: Propia

En la figura anterior se observa como los descriptores quedan orientados por la región de la que dependen, lo que permite independizar informaciones entre regiones.

5.3.3 Cálculo del histograma

El cálculo del histograma se corresponde con la generación del descriptor de SIFT. Este cálculo se lleva a cabo de manera que se logren conservar las propiedades de invariancia de escala, la invariancia de rotación conseguida con la orientación principal, invariancia ante cambios de iluminación, etc. Sin embargo, el objetivo de la modificación de esta etapa de SIFT tiene idéntico fin que la modificación de la etapa de cálculo de la orientación principal.

Se buscará anular la influencia en el histograma de la información fuera de la región del *descriptor activada* para cada *feature*. Junto con las *features* y sus regiones, se utilizará la orientación principal y demás parámetros que permitan llevar a cabo las normalizaciones requeridas.

Durante el proceso, se influirá de manera idéntica a como se hizo en el bloque de cálculo de orientación principal. En función de si los pixeles se encuentran o no en regiones activadas, se tratarán de una u otra manera tal y como se comentó en el apartado 5.3.1.

El resultado serán unos histogramas de magnitudes de gradiente que en total generarán vectores de valores que serán los descriptores asociados a cada *feature* de *RegSIFT*.

5.3.4 Generación del nuevo descriptor

Una vez se ha finalizado el procesado del algoritmo, es necesario mostrar los datos extraídos de manera que puedan ser utilizables en una aplicación. De esta misión se encargan los descriptores.

Se proponen dos descriptores distintos para este algoritmo:

- Descriptor de *features RegSIFT*:

Tal y como se comentó en el punto anterior 5.3.3 el descriptor básico consiste en la obtención y concatenación de cada uno de los valores del histograma de gradientes obtenido. Para mantener invariancias a cambios de luminosidad, sus valores se normalizan. Este proceso es en definitiva el que sigue SIFT pero con los nuevos datos. Cada descriptor obtenido corresponde con una *feature RegSIFT*. Como se demostrará en las pruebas de propiedades, sección 6.3, resulta menos robusto que su predecesor, aunque logra la corrección de los defectos de aplicación descritos en el Capítulo 4.

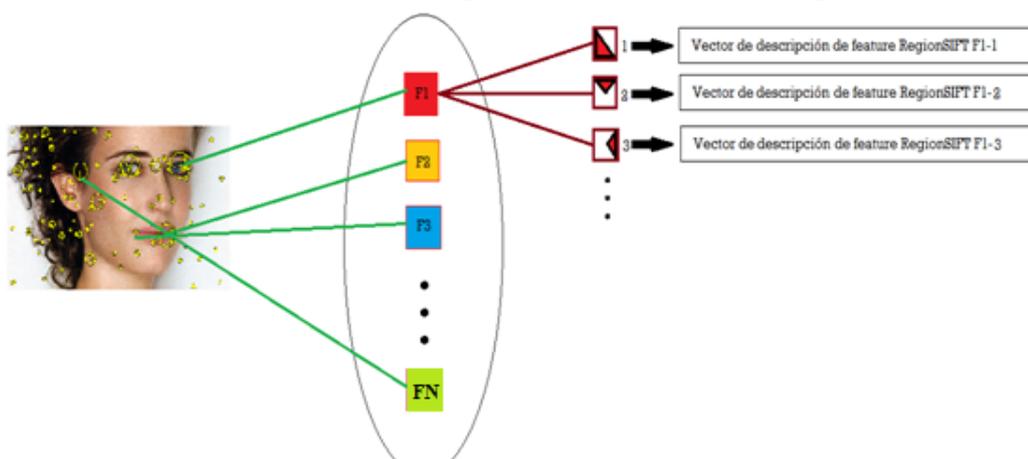


Figura 5-11: Esquema de descriptores individuales de *features RegSIFT*. Fuente: Propia

- Descriptor de *feature SIFT* original mediante *RegSIFT*:

La intención de este punto es la de mantener la detección que lleva a cabo SIFT, y describirla con el método *RegSIFT*.

Para llevar a cabo la generación del descriptor se sigue el siguiente proceso:

- i) Se extraen en primer lugar los descriptores de *features RegSIFT* que se han expuesto en el apartado anterior.
- ii) Se agrupan las *features RegSIFT*, esto es, sus descriptores, en cuanto a sus puntos SIFT de origen.
- iii) Se genera un descriptor de dicho punto SIFT de origen como la concatenación de los descriptores de *features RegSIFT* agrupados.

La longitud del descriptor será variable, por lo que la métrica que se deba aplicar para el cálculo de *matchings* deberá ser específica del método. Este descriptor será igual o más robusto que su predecesor (6.3) además de lograr la corrección de los defectos de aplicación. Esos descriptores se denominarán de aquí en adelante *groupedRegSIFT*.

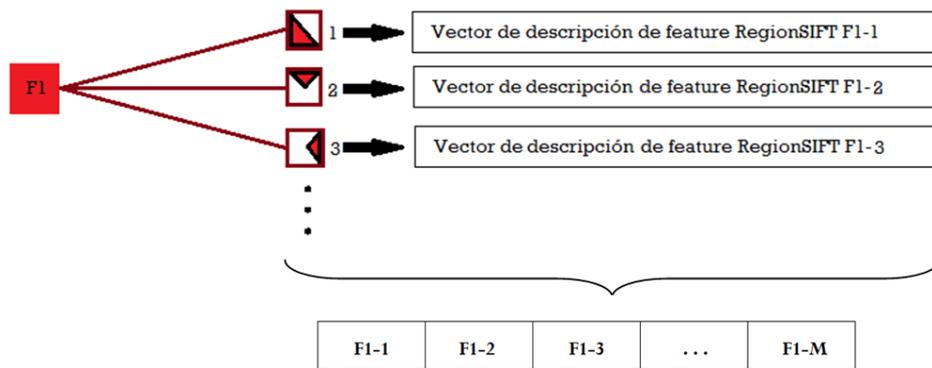


Figura 5-12: Esquema de conformación de descriptor de punto SIFT por *RegSIFT*.
Fuente: Propia

5.4 Resumen gráfico del algoritmo

Paso a paso pero de modo gráfico, se van a seguir los procesos llevados a cabo por el algoritmo sobre la imagen:

- 1) Detección de puntos SIFT y sus áreas de descripción



Figura 5-13: Imagen de análisis y detalle de detección SIFT con área de descripción superpuesta. **Fuente: Propia**

- 2) Segmentación de la imagen en *Superpíxeles*



Figura 5-14: Detalle del resultado de la segmentación en *Superpíxeles*. **Fuente: Propia**

3) División de las áreas de descripción según su pertenencia distintas regiones:



Figura 5-15: Áreas de descriptores de bordes divididas por regiones. Fuente: Propia

4) Extracción de los vectores de descripción según el método elegido:

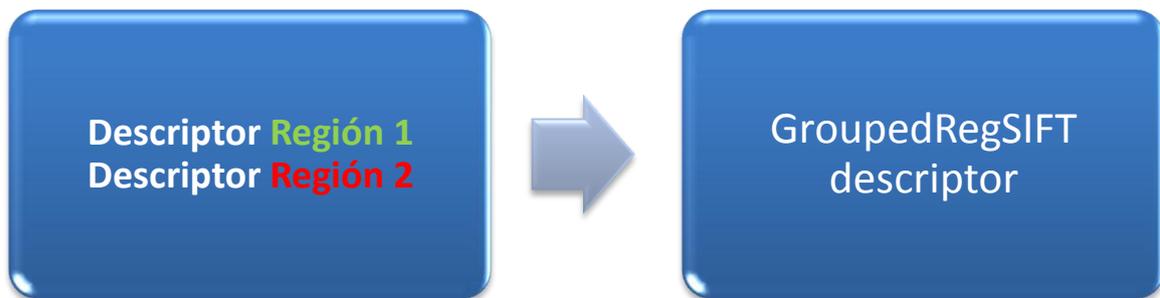


Figura 5-16: Obtención de los vectores por regiones, *RegSIFT* descriptor, o su conjunción *GroupedRegSIFT* descriptor. Fuente: Propia

5.5 Métricas asociadas

Tal y como se procede en el estado del arte, al enunciar un nuevo algoritmo, se incluye en su descripción una métrica recomendada por el autor para su evaluación sobre distintas aplicaciones o pruebas. La métrica asociada a los descriptores de este algoritmo, y de acuerdo a la cual se realizarán todas las pruebas a lo largo de los Capítulos 6 y 7 será la siguiente.

Para evaluar el comportamiento del algoritmo se realizarán comparaciones entre pares de conjuntos de descriptores SIFT y pares de conjuntos de los descriptores propuestos (tanto para *RegSIFT* como para *groupedRegSIFT*).

Sea:

\bar{f} un descriptor (o punto) SIFT (vector de 128 valores de longitud)

\bar{Rf} un descriptor *RegSIFT* (vector de 128 valores de longitud)

$\overline{GRf} = \{\bar{Rf}_1, \dots, \bar{Rf}_M\}$ un descriptor *groupedRegSIFT* (vector de $M \cdot 128$ valores de longitud)

Las comparaciones se realizarán de la siguiente manera.

5.5.1 Punto a punto

Dados dos puntos SIFT \bar{f}_1 y \bar{f}_2 , la distancia entre ambos se medirá mediante su norma L2:

$$D(\bar{f}_1, \bar{f}_2) = \|\bar{f}_1 - \bar{f}_2\|_2 = \sqrt{\sum_j (f_{1j} - f_{2j})^2} \quad \text{Ecuación 10}$$

De igual manera, dados dos puntos *RegSIFT* \overline{Rf}_1 y \overline{Rf}_2 :

$$D(\overline{Rf}_1, \overline{Rf}_2) = \|\overline{Rf}_1 - \overline{Rf}_2\|_2 = \sqrt{\sum_j (Rf_{1j} - Rf_{2j})^2} \quad \text{Ecuación 11}$$

Para el caso de dos descriptores *groupedRegSIFT*: $\overline{GRf}^1 = \{\overline{Rf}_1^1 \dots \overline{Rf}_{M1}^1\}$ y $\overline{GRf}^2 = \{\overline{Rf}_1^2 \dots \overline{Rf}_{M2}^2\}$, la distancia se calculará simplemente como el mínimo entre todas las comparaciones *RegSIFT* posibles:

$$D(\overline{GRf}^1, \overline{GRf}^2) = \min \left(\|\overline{Rf}_{m1}^1 - \overline{Rf}_{m2}^2\|_2 \right), \quad \begin{cases} \forall m1 \in [1, M1] \\ \forall m2 \in [1, M2] \\ m1, m2 \in \mathbb{N} \end{cases} \quad \text{Ecuación 12}$$

5.5.2 Punto a conjunto de puntos

Sea un punto SIFT \bar{f} y un conjunto de puntos SIFT $F = \{\bar{f}_1 \dots \bar{f}_N\}$, $N = |F|$, la distancia del punto al conjunto corresponderá al mínimo de las distancias obtenidas entre el punto SIFT y cada uno de los puntos dentro del conjunto, es decir, abusando de la nomenclatura:

$$D(\bar{f}, F) = \min (D(\bar{f}, \bar{f}_n)), \quad \forall n \in [1, N], n \in \mathbb{N} \quad \text{Ecuación 13}$$

Igualmente, sea $RF = \{\overline{Rf}_1 \dots \overline{Rf}_N\}$, $N = |RF|$ un conjunto de descriptores *RegSIFT*:

$$D(\overline{Rf}, RF) = \min (D(\overline{Rf}, \overline{Rf}_n)), \quad \forall n \in [1, N], n \in \mathbb{N} \quad \text{Ecuación 14}$$

Finalmente, sea \overline{GRf} un descriptor *groupedRegSIFT* y $GRF = \{\overline{GRf}^1 \dots \overline{GRf}^N\}$, $N = |GRF|$, un conjunto de descriptores *groupedRegSIFT*:

$$D(\overline{GRf}, GRF) = \min (D(\overline{GRf}, \overline{GRf}^n)), \quad \forall n \in [1, N], n \in \mathbb{N} \quad \text{Ecuación 15}$$

5.5.3 Conjunto de puntos a conjunto de puntos

Para cada descriptor SIFT, *RegSIFT* y *groupedRegSIFT* de uno de los conjuntos (de búsqueda o de *train*) se calculará la distancia al otro conjunto (buscado o de *test*) según lo indicado en las ecuaciones (Ecuación 10-Ecuación 15). La relación se llevará a cabo desde los elementos de *train* a todos los elementos de *test* sin repetición, lo que dará lugar a una relación inyectiva (el proceso está ilustrado en la Figura 5-17)

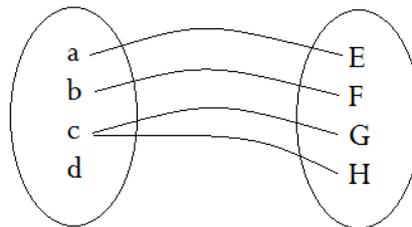


Figura 5-17: Relación inyectiva. Fuente: Propia

Capítulo 6 – Evaluación del Algoritmo

6.1 Introducción

En este capítulo se procederá a chequear los comportamientos y propiedades logradas con el nuevo método *RegSIFT*.

En todas las pruebas realizadas, se fijará el valor del umbral de replicación por región $Oth = 33,3\%$, ver sección 5.3.1. También referenciar que de aquí en adelante, se hablará de las regiones resultantes del segmentador, *Superpíxeles*, como *SP*.

En primer lugar, sobre un *toy-example* se comprobará que se han conseguido solucionar los defectos de aplicación expuestos en el Capítulo 4. Se evaluarán los comportamientos del método propio y de SIFT en una serie de situaciones que permitan observar dichos defectos y se observará si han sido solventados por *RegSIFT*.

En segundo lugar, y previo a su aplicación en distintas tareas, se comprobará que las propiedades que hacen de SIFT el método más referenciado para la descripción de imágenes se mantienen en el nuevo algoritmo.

6.2 Toy-Example

6.2.1 Introducción

Se ha desarrollado un sistema de evaluación de tipo *toy-example* que permita comparar en un entorno de pruebas controlado el método de SIFT original con el propuesto.

Más concretamente, se va a estudiar el comportamiento de los puntos SIFT en las regiones de borde, donde se generan los defectos por introducir información del entorno no deseada, frente al comportamiento del algoritmo desarrollado en esas mismas zonas.

6.2.2 Metodología de evaluación

6.2.2.1 Dataset

Se ha creado una serie de imágenes artificiales para la evaluación. Dichas imágenes se pueden catalogar en dos grupos:

- Fondos *Poco* Texturados: 4 Imágenes (I01, I02, I03 e I04).
- Fondos *Altamente* Texturados: 4 Imágenes (IT01, IT02, IT03 e IT04).

Adicionalmente se dispone también de las máscaras de *ground-truth* que permitirán realizar la clasificación de los puntos para centrar el estudio, tal y como se mencionó en la introducción, en los problemáticos puntos de borde.

Si bien los montajes de los objetos con los fondos han sido propios, se ha recurrido unos cromas con las citadas máscaras de *ground-truth* obtenidos de una base de datos utilizada en el artículo [44]. Todas las imágenes, máscaras y cromas se encuentran en el A.A. A continuación se muestran dos ejemplos de cada fondo.



Figura 6-1: I01 e IT01. Fuente: Propia y [44]

6.2.2.2 Clases

Sobre las imágenes del *dataset* se van a ejecutar los algoritmos a evaluar, esto es, SIFT y *RegSIFT*.⁶

En función de la ubicación sobre la máscara de frente-fondo, *ground-truth*, de las distintas detecciones que lleve a cabo cada método, se van a clasificar los puntos en diversas clases, definidas como F (*foreground*-frente), B (*background*-fondo) y E (*edge*-borde).

SIFT	F	B	E
<i>RegSIFT</i>	F	B	EF –EB

Tabla 6-1: Clases de puntos *toy-example*

Dado que los puntos de *RegSIFT* van a utilizar distintas regiones del descriptor SIFT para describir el punto, los puntos localizados en el borde se clasificarán de este modo según la región del descriptor utilizada para describir esa *feature RegSIFT* sea interna o externa al objeto. No aplica en este caso la utilización de un umbral para distinguir entre EF y EB, dado que la naturaleza sintética de las imágenes utilizadas ha conllevado la aparición de regiones en exclusivamente alguna de las dos clases, nunca intermedias.

6.2.2.3 Métrica

Para la extracción de resultados y estadísticos, se seguirá un proceso que consistirá en los siguientes pasos:

- 1) *Matching* de puntos
Los *matchings* de los grupos de puntos de cada imagen se realizarán según lo expuesto en la sección 5.5.2.
- 2) Clasificación de los *matchings*
Una vez realizados los *matchings* de puntos con sus distancias asociadas, se procederá a umbralizar las mismas.
Como resultado de la umbralización y de las clases fijadas en la sección 6.2.2.2 aparecerán los siguientes tipos de asociaciones:

⁶ Nótese que solo se incluirán resultados con los descriptores *RegSIFT*, menos robustos que los *groupedRegSIFT*, como se verá en la sección 6.3, lo que hará que la prueba sea más exigente.

RegSIFT	Distancia \leq Umbral	Distancia $>$ Umbral
F \leftrightarrow F	TruePositives (TP)	FalseNegatives (FN)
F \leftrightarrow EF	TruePositives (TP)	FalseNegatives (FN)
EF \leftrightarrow EF	TruePositives (TP)	FalseNegatives (FN)
B \leftrightarrow F	FalsePositives (FP)	TrueNegatives (TN)
B \leftrightarrow EF	FalsePositives (FP)	TrueNegatives (TN)
EB \leftrightarrow F(*)	FalsePositives (FP)	TrueNegatives (TN)
EB \leftrightarrow EF(*)	FalsePositives (FP)	TrueNegatives (TN)

Tabla 6-2: Métrica RegSIFT toy-example Fuente: Propia

(*) Las *features* de EB se han clasificado como de fondo puesto que las regiones de las que extraen finalmente la información son externas al objeto, por lo que de ninguna manera se podrían considerar TruePositives *matchings* de información del frente con esta información de fondo. Carece pues de toda lógica la búsqueda de *matchings* $X \leftrightarrow B$ puesto que los fondos varían entre una imagen y otra.

SIFT	Distancia $<$ Umbral	Distancia $<$ Umbral
F \leftrightarrow F	TP	FN
F \leftrightarrow E	TP	FN
E \leftrightarrow E	TP	FN
B \leftrightarrow F	FP	TN
B \leftrightarrow E	FP	TN

Tabla 6-3: Métrica SIFT toy-example Fuente: Propia

A partir de este punto, resulta necesario destacar que, puesto que los comportamientos de SIFT en el frente del objeto no se han detectado como erróneos, se ceñirá el resto del análisis a las *features* detectadas en los bordes.

6.2.2.4 Estadísticos

Una vez fijadas las clases (6.2.2.2), la métrica (6.2.2.3) y las zonas de evaluación (bordes), se va a proceder a establecer los estadísticos que se van a extraer como resultados del proceso.

- **TP/FP** - Mediante un barrido de umbrales, se observará la evolución de estas dos métricas.
- **Precisión/Recall** – De nuevo barriendo umbrales de distancias, se observará cómo evoluciona la relación entre los TP-FN y de los FP.

$$Recall = \frac{\#TP}{\#TP + \#FN} \quad \text{Ecuación 16}$$

$$Precisión = \frac{\#TP}{\#TP + \#FP} \quad \text{Ecuación 17}$$

- **FScore** – En función de los parámetros anteriores, Precisión y *Recall* se mostrará un valor conjunto de ambos que de una impresión del comportamiento del sistema a nivel general.

$$FScore = 2 * \frac{Precisión * Recall}{Precisión + Recall} \quad \text{Ecuación 18}$$

6.2.2.5 Objetivo de la evaluación

Los objetivos que se persiguen con esta evaluación no han sido otros que los siguientes:

- Comprobar el comportamiento general del método *RegSIFT* en los bordes de un objeto (frente) replicado sobre fondos diversos con distintos grados de información.
- Comprobar que la distribución de las distancias de los *matchings* positivos (TP) del método propio es inferior, y por tato mejor, que la de SIFT básico.
- Comprobar que las distancias de los *matchings* negativos (FP) mejoran la relación con respecto a las distancias de los *matchings* positivos, esto es, crecen en mayor medida. Como veremos en la sección 6.2.3, esta mejoría será más notable en fondos altamente texturados.

Al satisfacerse todos estos objetivos, se puede concluir que, el método desarrollado mejora el funcionamiento de SIFT en el objetivo de la descripción de objeto con independencia de sus fondos.

6.2.3 Resultados

Los resultados obtenidos no serán otros que los datos para cada uno de los estadísticos presentados en la sección 6.2.2.4. Cabe referenciar que los valores utilizados para las representaciones gráficas se han obtenido de las medias de los valores para cada comparación por pares de las imágenes del *dataset*.

Fondos Poco Texturados

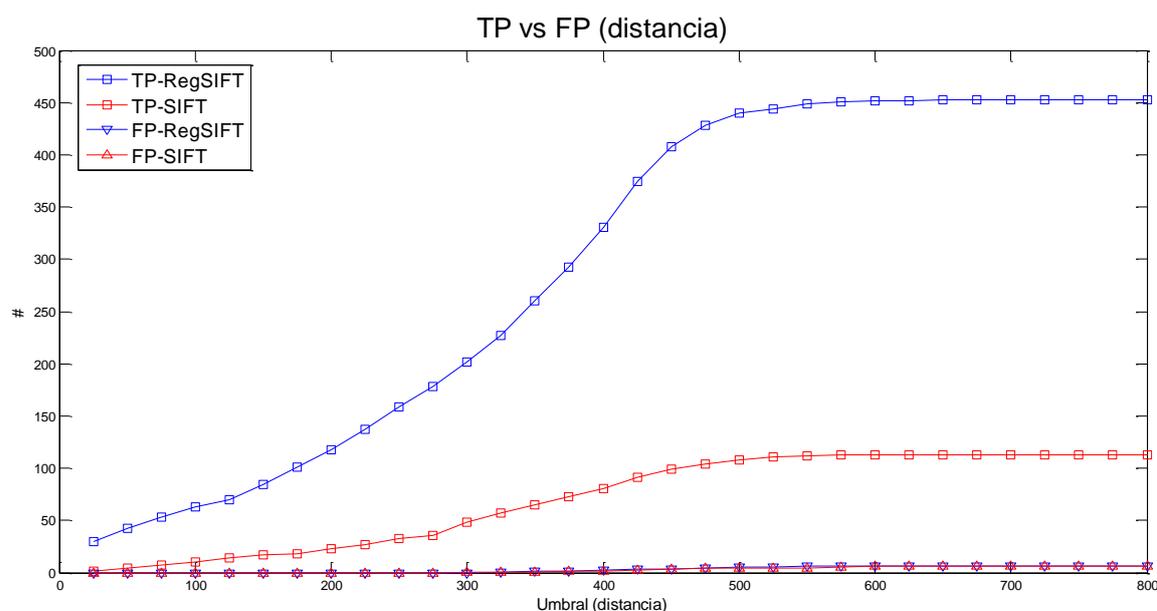


Figura 6-2: TP / FP para Fondos Poco Texturados.

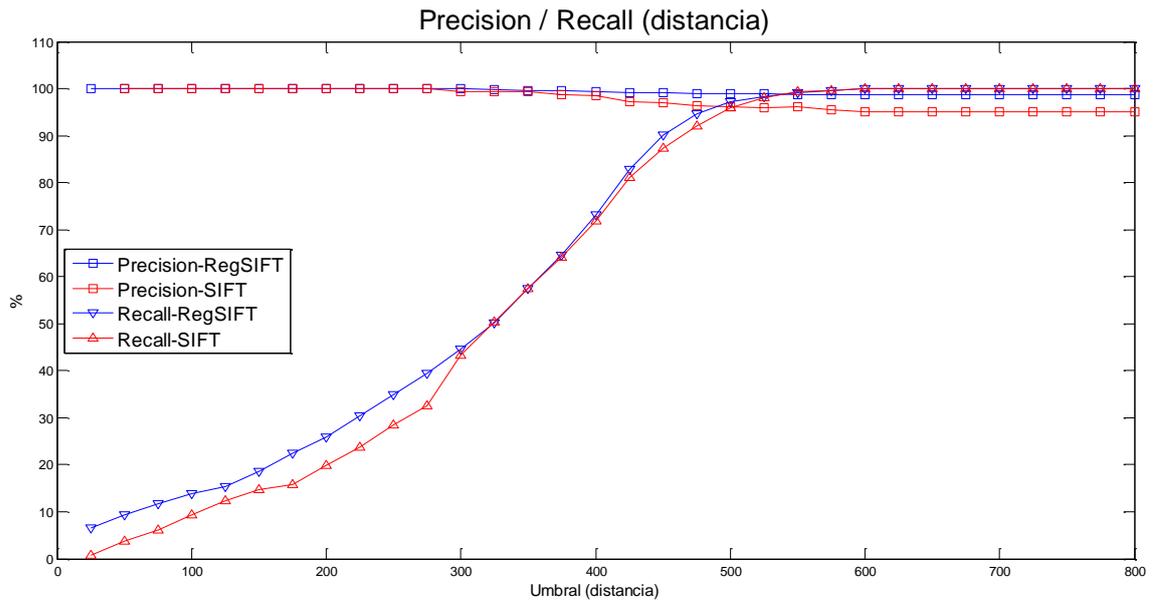


Figura 6-3: Precisión/Recall para Fondos Poco Texturados.

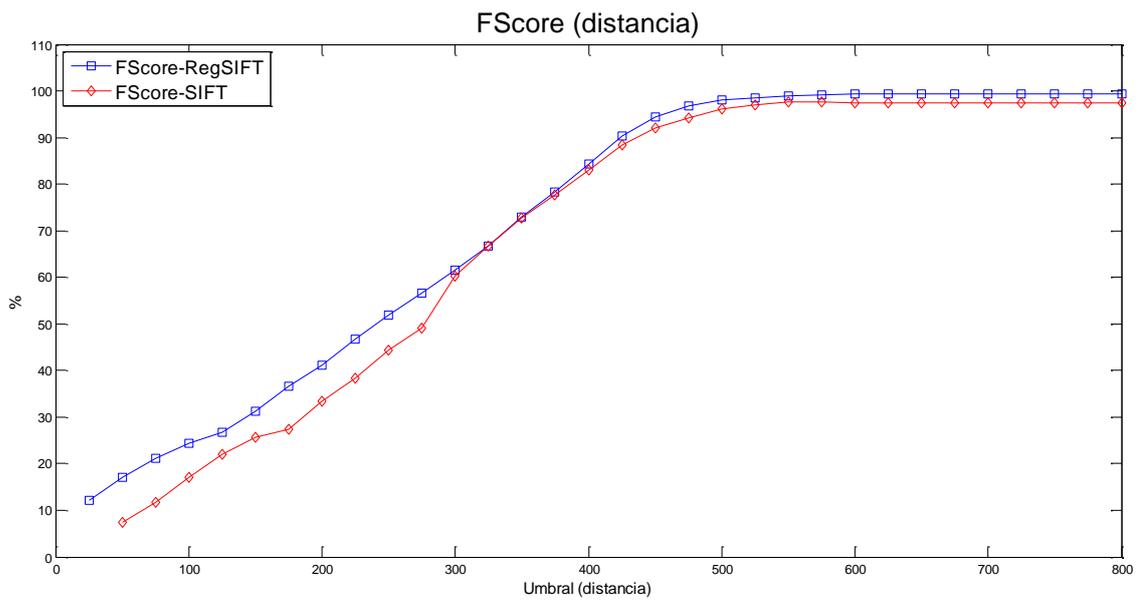


Figura 6-4: FScore para Fondos Poco Texturados.

Fondos Altamente Texturados

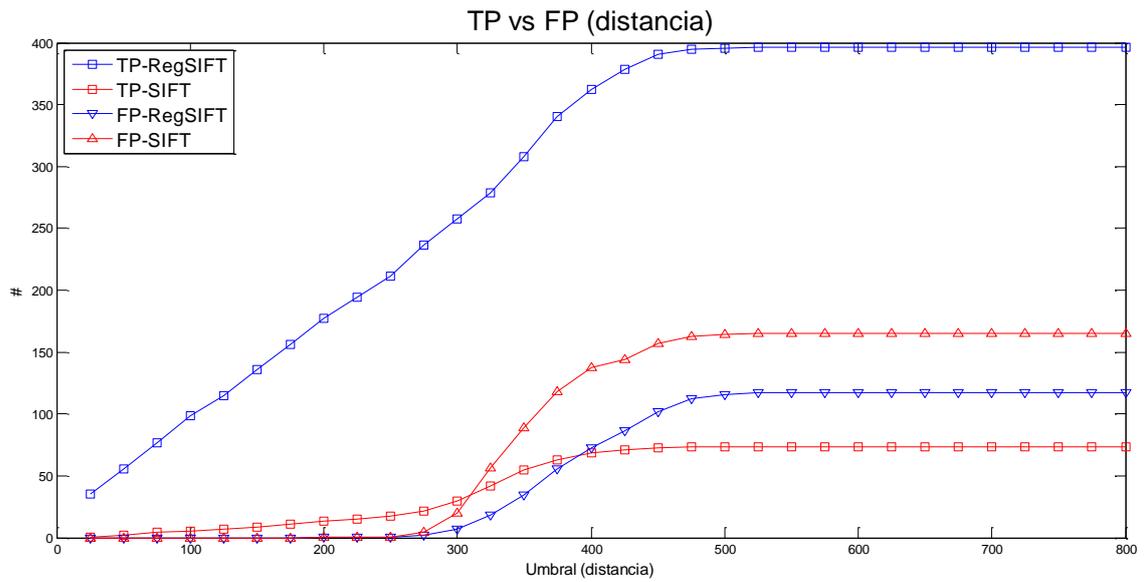


Figura 6-5: TP / FP para Fondos Altamente Texturados.

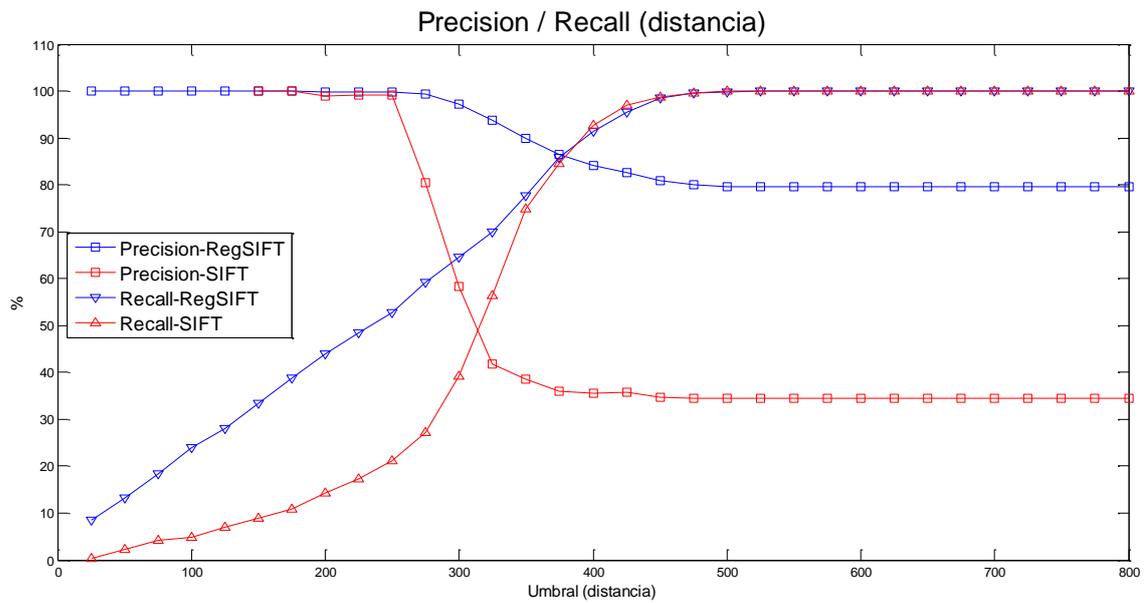


Figura 6-6: Precisión/Recall para Fondos Altamente Texturados.

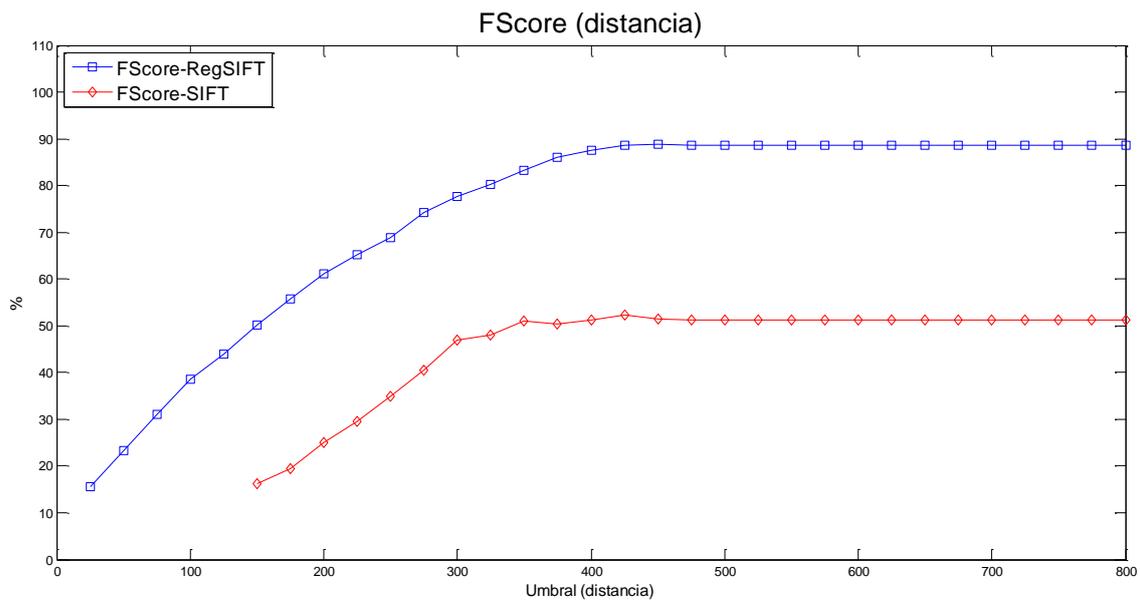


Figura 6-7: FScore para Fondos Altamente Texturados.

6.2.4 Análisis de resultados

Para ser exhaustivo en el estudio de los resultados obtenidos, se procederá a estudiar los mismos según los dos tipos de imágenes, Fondos Poco Texturados y Fondos Altamente Texturados.

En primer lugar se estudiarán los resultados obtenidos sobre las imágenes con Fondos Poco Texturados, esto es, Imágenes I01, I02, I03 e I04 (ver Anexo A.1).

En segundo lugar se analizarán los datos resultantes de las pruebas con imágenes de Fondos Altamente Texturados, IT01, IT02, IT03 e IT04 (ver Anexo A.1).

Fondos Poco Texturados

Precisión:

- Debido a lo variable de los descriptores SIFT en los bordes, la precisión es muy elevada, ya que se producen muy pocos FP.
- Igualmente, los valores de *RegSIFT* son también muy altos en la precisión. Las *features* nuevas parecen resultar lo suficientemente distintivas.
- Hasta umbralizar con distancias medias-altas, la Precisión se mantiene en torno a 1, lo que permite suponer que las distancias de los FP serán altas, tal y como se desea.

Recall:

- Comportamiento superior del método propio en los primeros umbrales. Esto permite presuponer que los valores de distancias entre TP son bajas.
- Punto de alcance de *Recall* = 100% similar para ambos métodos en este tipo de fondos.

FScore:

- Pese a que el entorno resulta muy favorable para el método SIFT, el método propio consigue superar sus resultados pese al escaso margen.
- Se alcanza un FScore de casi 100% en ambos métodos, pese a que se deba a distintos motivos.

Los parámetros evaluados, junto con la gráfica de #TP y #FP, recordando que se habla siempre de puntos de borde, permiten extraer dos conclusiones.

- SIFT, por sus características realiza pocos *matchings* TP en los bordes debido a los defectos comentados en el Capítulo 4. Sin embargo, dos fenómenos palián este efecto nocivo en este tipo de fondos. En primer lugar, con fondos poco texturados, el número de puntos detectados en los bordes es reducido en proporción al total. En segundo lugar, puesto que estos puntos son descritos de manera errónea, el número de FP es también bajo, por lo que los defectos en estos casos no resultan tan nocivos estadísticamente, mientras que si lo serán cuando los fondos sean altamente texturados.
- Por otra parte, el incremento del número de TP en *RegSIFT* a bajas distancias distanciándolos de los potenciales FP, permite concluir que el nuevo método implementado consigue generar en las *features* de la imagen un efecto que se podría denominar como de “separación de Gaussianas”, mostrado y expuesto a continuación.

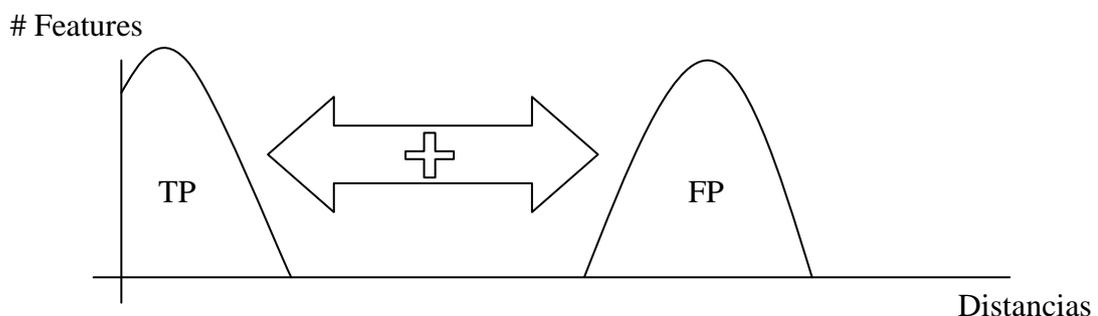


Figura 6-8: Efecto de separación de Gaussianas. Fuente: Propia

El efecto de “separación de Gaussianas” supone que, los puntos potenciales de ser FP incrementan o mantienen sus distancias, mientras que los que generan TP las reducen. Este efecto permitirá diseñar estrategias de umbralización más flexibles, incluyendo menor número de errores al fijar un umbral.

Fondos texturados

Se trata de una situación mucho más compleja que la anterior. En esta situación, SIFT se presupone que ha de empeorar su comportamiento, principalmente en los bordes. Los resultados obtenidos se pueden resumir como:

- El número de FP aumenta en ambos casos, por lo que la precisión empeora. Sin embargo, gracias al efecto de “separación de Gaussianas” citado anteriormente, la Precisión del método propio acaba quedando por encima de los valores de SIFT.
- El *Recall* empeora muy notablemente en ambos casos debido al gran número de *features* detectadas en los bordes que realmente no se deben al propio borde del objeto de frente sino a las texturas colindantes. Sin embargo, el buscado efecto de independizar el frente en la descripción hace que el *Recall*, principalmente en los primeros umbrales sea superior.
- Estas dos situaciones redundan en unos valores de FScore netamente muy superiores. En el punto óptimo de ambas para los FScore, la diferencia del método propio al de SIFT se encuentra en torno al 40%, ver Figura 6-7.

Adicionalmente, observando las gráficas de # de TP y FP se concluye que:

- Para los TP, las distancias se han mantenido prácticamente estables, mientras que las de SIFT se han disparado, por lo que el método da un resultado notablemente mejor.
- Para los FP el efecto es aun más beneficioso, pues las distancias obtenidas para estos son siempre superiores, y por lo tanto más diferentes.

En definitiva, se puede decir en base a este *toy-example*, que se han logrado corregir los defectos inherentes a SIFT en lo que a incorporación de información del entorno no deseada se refiere.

Resta evaluar la conservación de las propiedades del algoritmo original y la utilización del método en aplicación real.

6.3 Evaluación de propiedades

6.3.1. Introducción

El objetivo de este proyecto no es otro sino el de mejorar el funcionamiento de un método objetivo del estado del arte del procesamiento de imágenes. Pese a no haberse especificado concretamente, la mejora de alguno de los definidos como defectos de aplicación (Capítulo 4) debe hacerse manteniendo intacta las propiedades del método que no se encuentren directamente ligadas con los aspectos cambiados.

Con el fin de comprobar que se cumple dicha premisa se ha desarrollado este apartado de evaluación de propiedades, en el que se evaluarán concretamente las propiedades presentadas por SIFT expuestas en la sección 3.3

Para la evaluación, se recurrirá al *dataset* de imágenes empleando por Mikolajczyk en su artículo comparativo de descriptores [13]. Sobre él se extraerán los resultados pertinentes de SIFT y *RegSIFT* y se comentarán los resultados.

Las propiedades evaluadas serán:

- Invariancia a rotación.
- Invariancia a los cambios de escala.
- Invariancia a cambios de iluminación.
- Invariancia a borrosidad.
- Invariancia a cambios del punto de vista.

6.3.2. Metodología de evaluación

6.3.2.1 Dataset

Tal y como se ha expuesto en la sección anterior 6.3.1, se va a emplear un *dataset* libre empleado en el artículo comparativo [14].

El *dataset* consta de 24 imágenes, ver Anexo A.B clasificadas en 4 familias según las propiedades que es posible evaluar sobre ellas. Las familias por propiedades son:



Figura 6-9: Ejemplos del dataset por familias de propiedades. Fuente: [13]

6.3.2.2 Métrica

Las pruebas llevadas a cabo se han realizado siguiendo el modelo de *matching* de *features* expuesto en la sección 5.5.3. Esta métrica es idéntica a la empleada en el artículo de origen del *dataset* [14] salvo porque este añade a las distancias una búsqueda por *nearest-neighbour* (NN), que facilitan los aciertos, y por lo que no se usará.

Una vez fijados los *matchings*, se procederá a evaluar las distancias obtenidas para cada par de imágenes de la misma familia de propiedad. Las distribuciones de distancias, y principalmente su evolución con la variación de las propiedades, esto es, incremento del *blur*, cambios del punto de vista, incrementos de la rotación,..., serán evaluados para verificar o no la conservación de las propiedades deseadas del método original.

En este apartado de métrica, también cabe destacar que las pruebas se realizarán esta vez utilizando los dos tipos de descriptores propuestos comentados en la sección 5.3.4.

6.3.3. Resultados

Los resultados se presentan a continuación clasificados por familias. Dentro de cada familia, se mostrarán gráficamente las distribuciones de distancias de la imagen original a cada una de las de dicha categoría. Se compararan las distribuciones para SIFT y los descriptores de *RegSIFT* y *groupedRegSIFT*.

Para obtener las distribuciones de distancias se ha utilizado una función para estimación de la función de densidad desarrollada por Bowman en [45] con kernel Gaussiano.

6.3.3.1 Blur

El conjunto de imágenes utilizado ha sido:



Figura 6-10: Dataset de imágenes de blur. De menor a mayor blur (izda. a dcha. y de arriba a abajo). Fuente: [14]

Los resultados obtenidos:

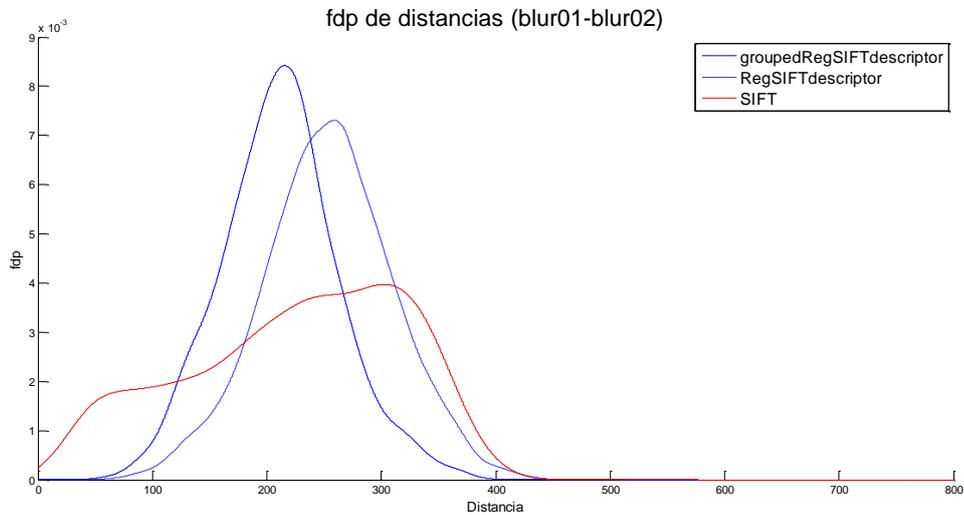


Figura 6-11: fdp de distancias para blur01-blur02

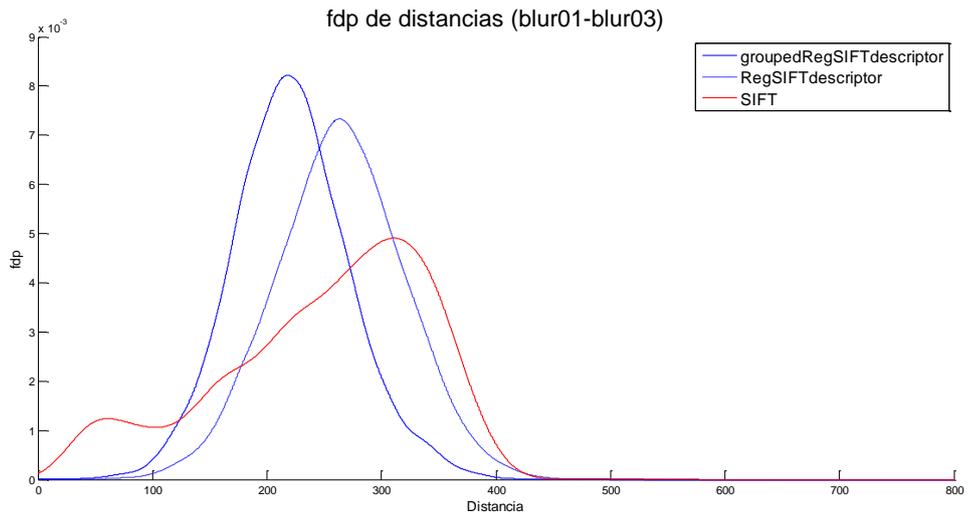


Figura 6-12: fdp de distancias para blur01-blur03

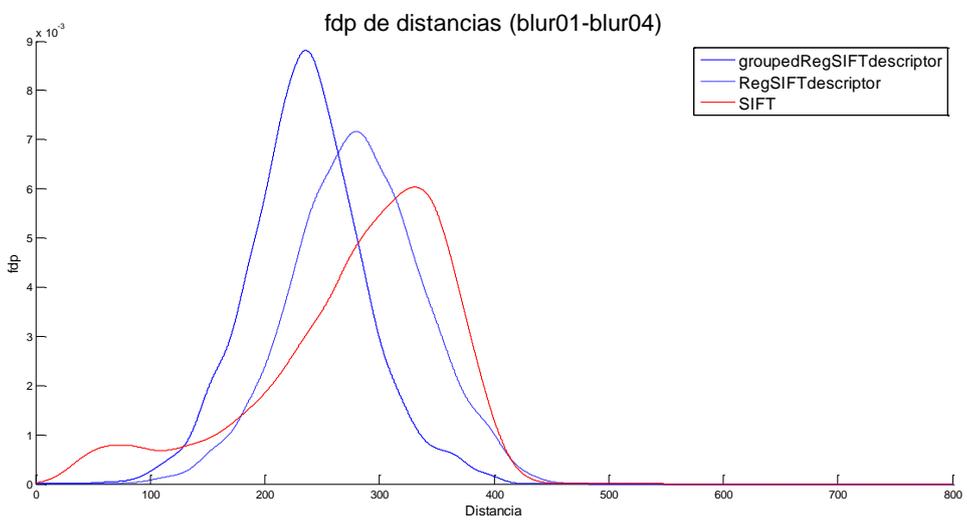


Figura 6-13: fdp de distancias para blur01-blur04

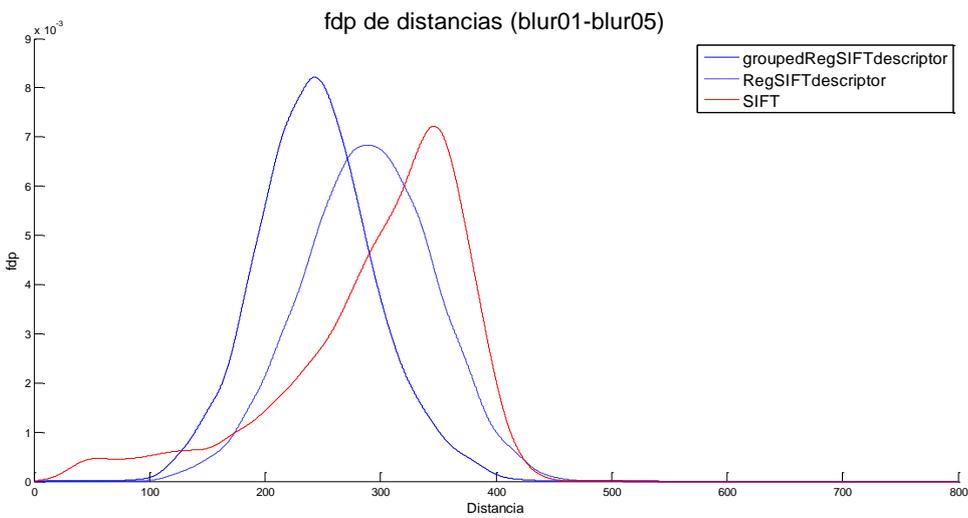


Figura 6-14: fdp de distancias para blur01-blur06

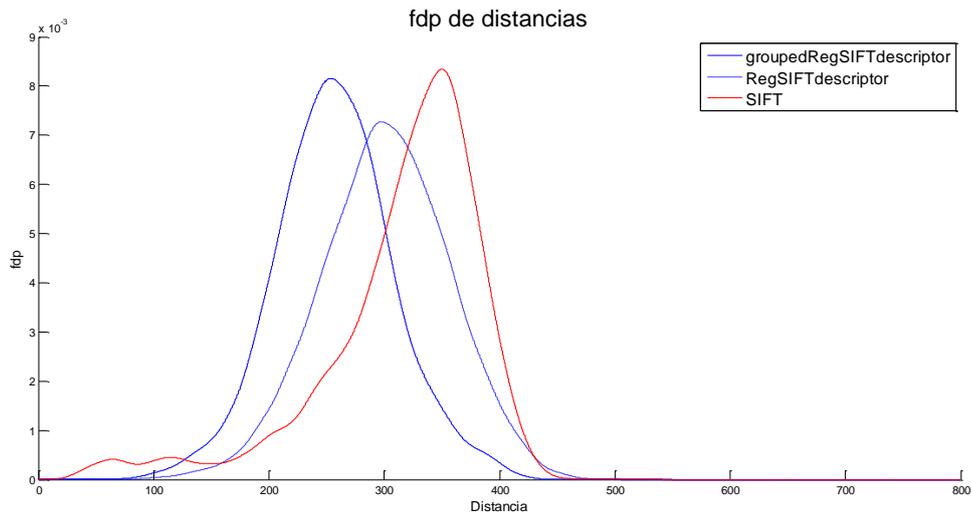


Figura 6-15: fdp de distancias para blur01-blur06

6.3.3.2 Light

El conjunto de imágenes utilizado ha sido:



Figura 6-16: Dataset de imágenes de *light*. De menor a mayor *light* (izda. a dcha. y de arriba a abajo). Fuente: [14]

Los resultados obtenidos:

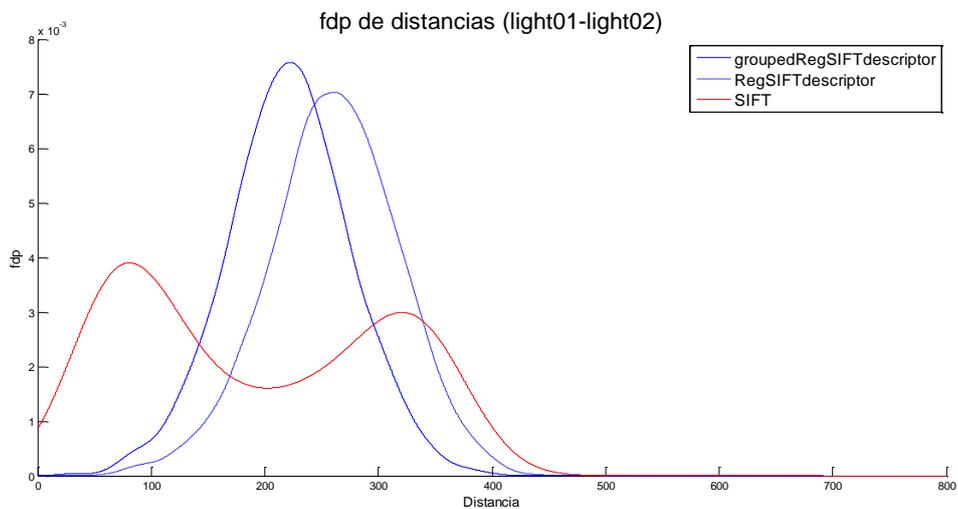


Figura 6-17: fdp de distancias para light01- light 02

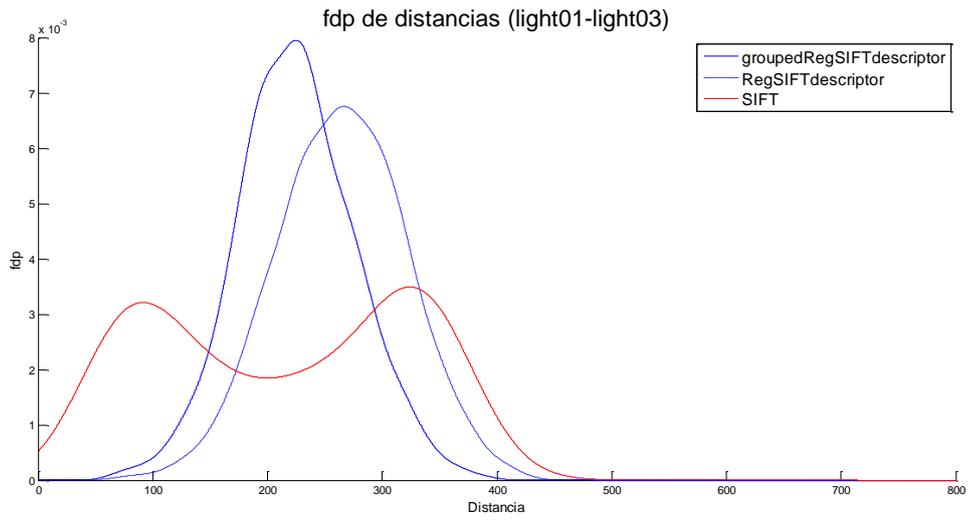


Figura 6-18: fdp de distancias para light01- light 03

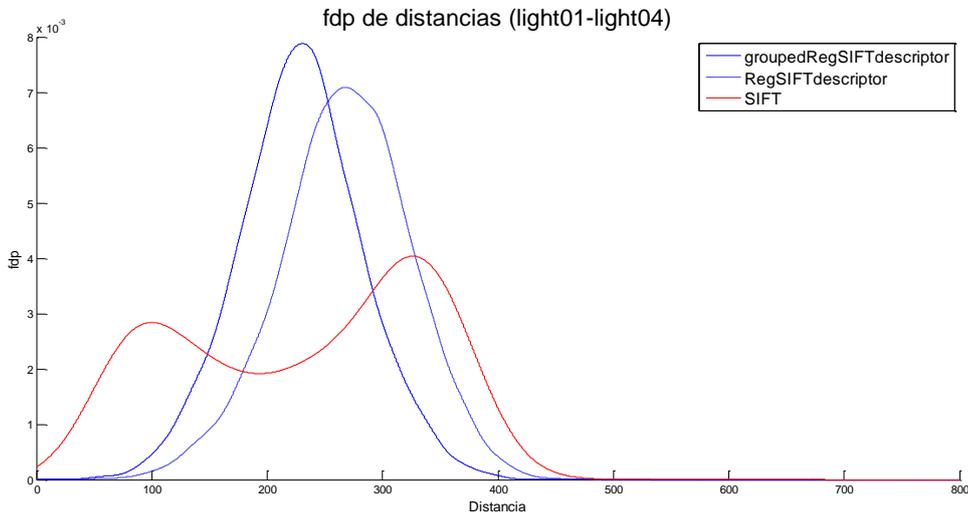


Figura 6-19: fdp de distancias para light01- light 04

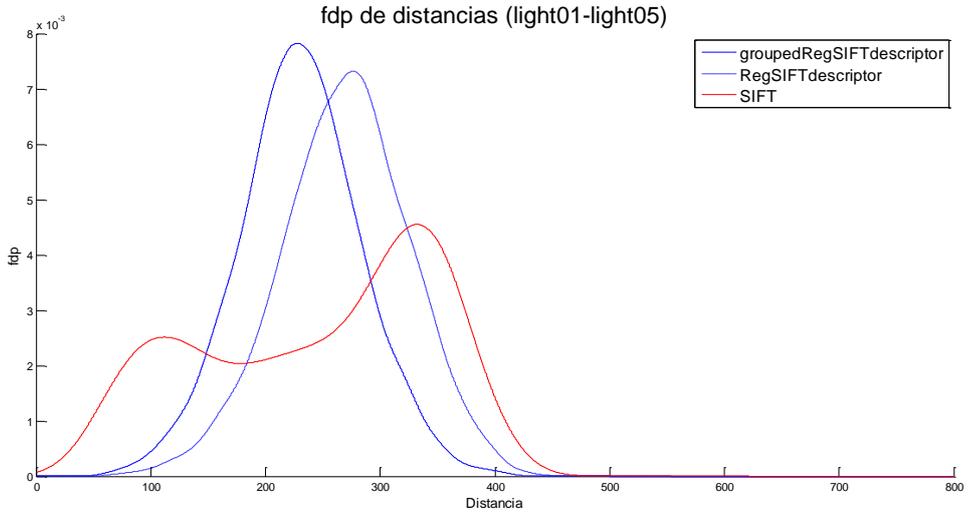


Figura 6-20: fdp de distancias para light01- light 05

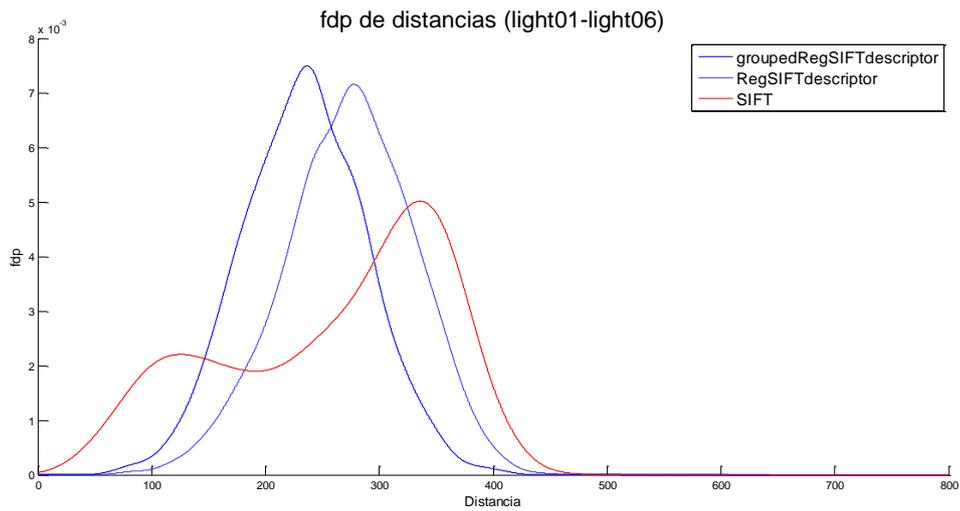


Figura 6-21: fdp de distancias para light01- light 06

6.3.3.3 Viewpoint

El conjunto de imágenes utilizado ha sido



Figura 6-22: Dataset de imágenes de viewpoint. De menor a mayor ángulo de viewpoint (izda. a dcha. y de arriba a abajo). Fuente: [14]

Los resultados obtenidos:

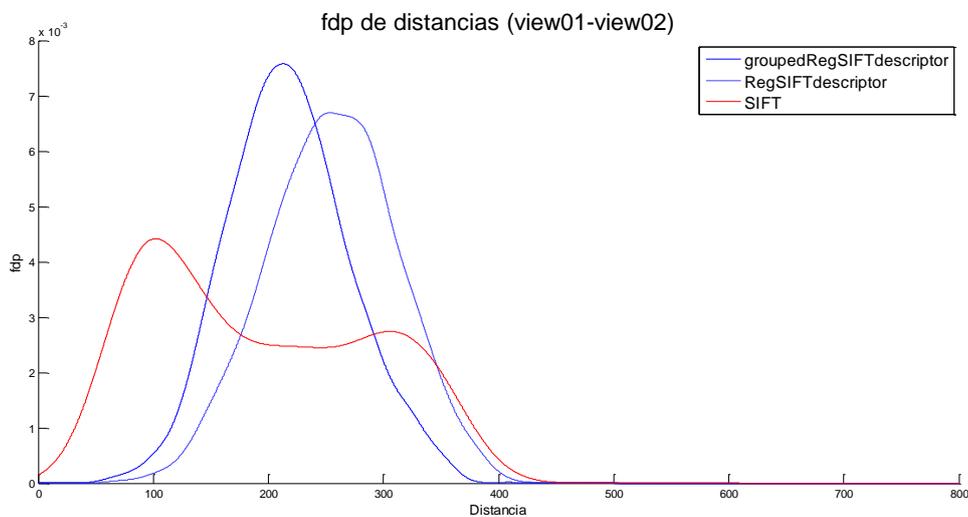


Figura 6-23: fdp de distancias para viewpoint01- viewpoint02

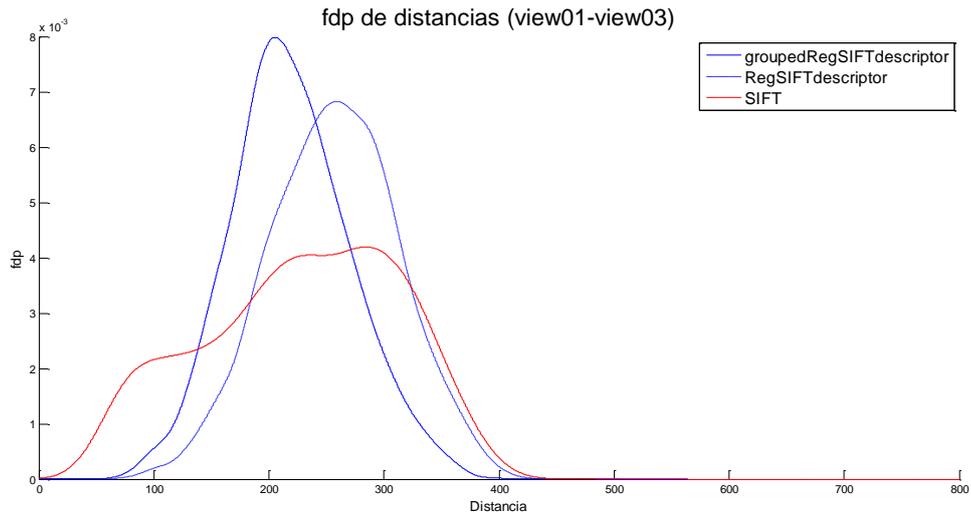


Figura 6-24: fdp de distancias para viewpoint01- viewpoint03

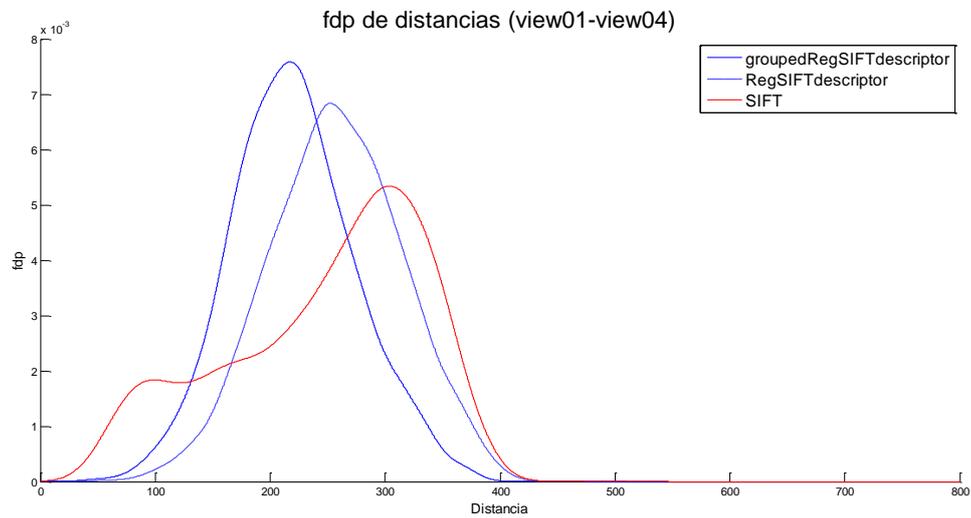


Figura 6-25: fdp de distancias para viewpoint01- viewpoint04

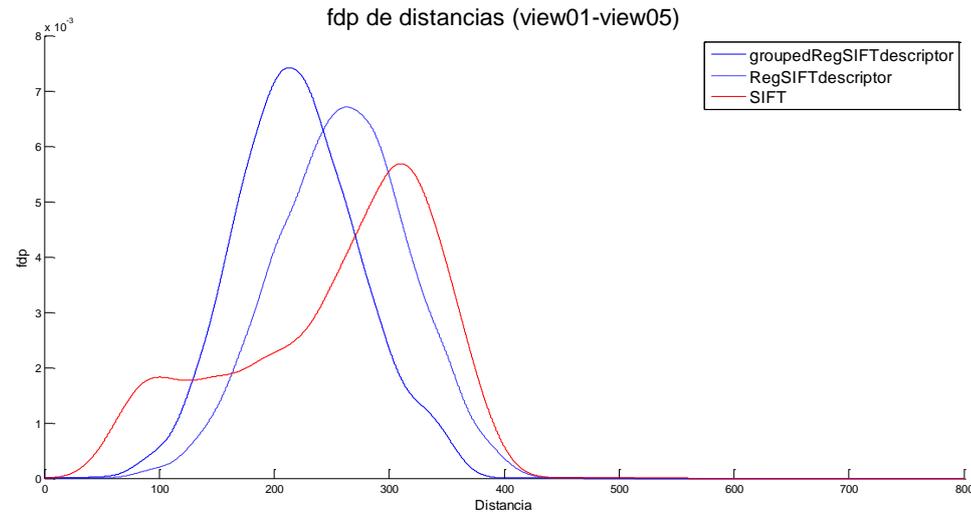


Figura 6-26: fdp de distancias para viewpoint01- viewpoint05

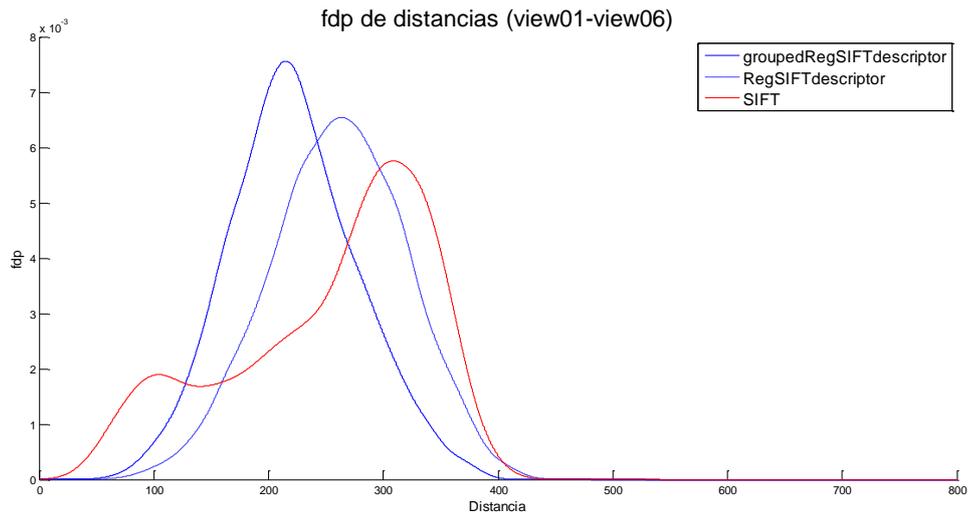


Figura 6-27: fdp de distancias para viewpoint01- viewpoint06

6.3.3.4 Scale + Rotate

El conjunto de imágenes utilizado ha sido:



Figura 6-28: Dataset de imágenes de scale+rotate. Fuente: [14]

Los resultados obtenidos:

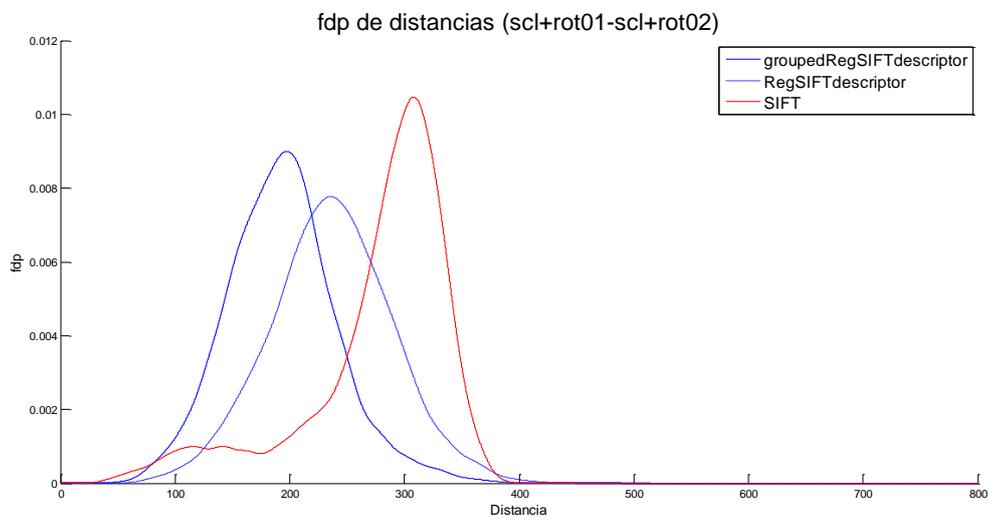


Figura 6-29: fdp de distancias para scl+rot01- scl+rot02

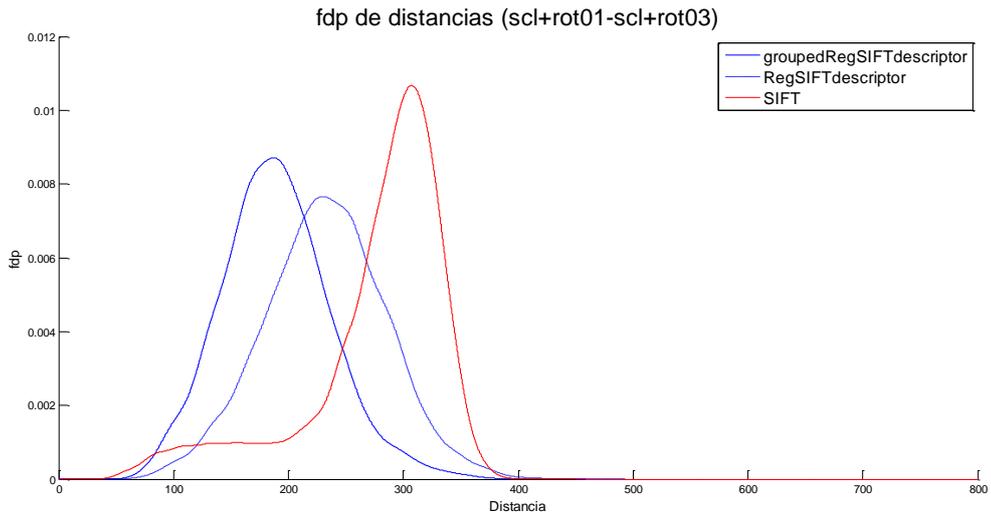


Figura 6-30: fdp de distancias para scl+rot01- scl+rot03

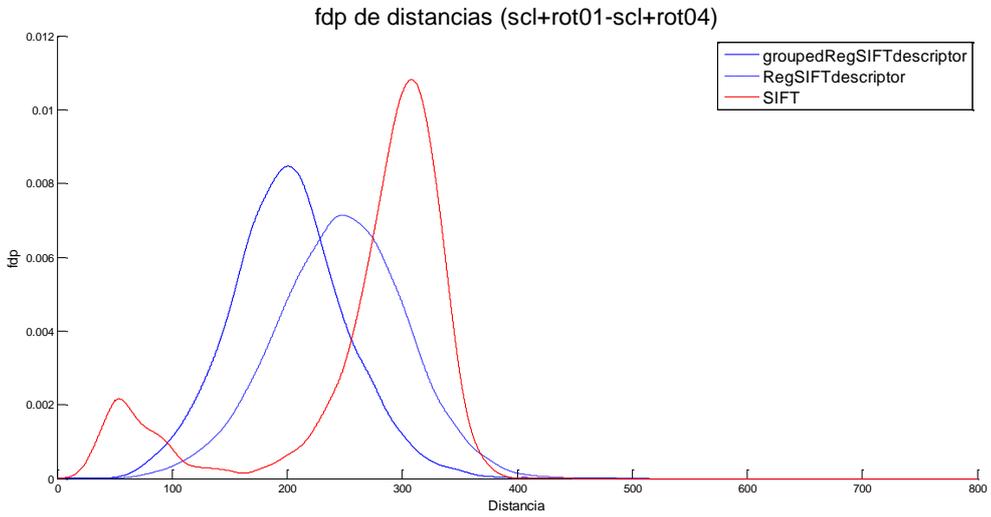


Figura 6-31: fdp de distancias para scl+rot01- scl+rot04

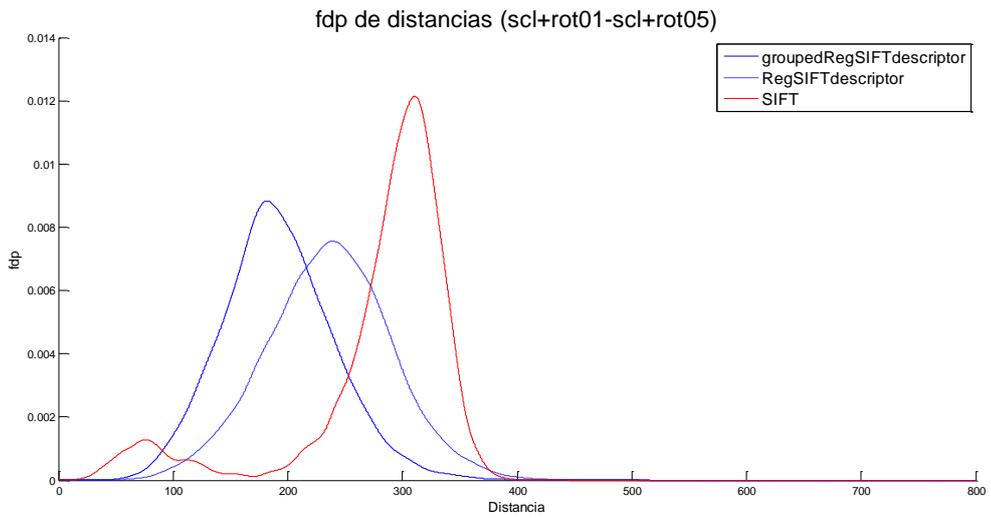


Figura 6-32: fdp de distancias para scl+rot01- scl+rot05

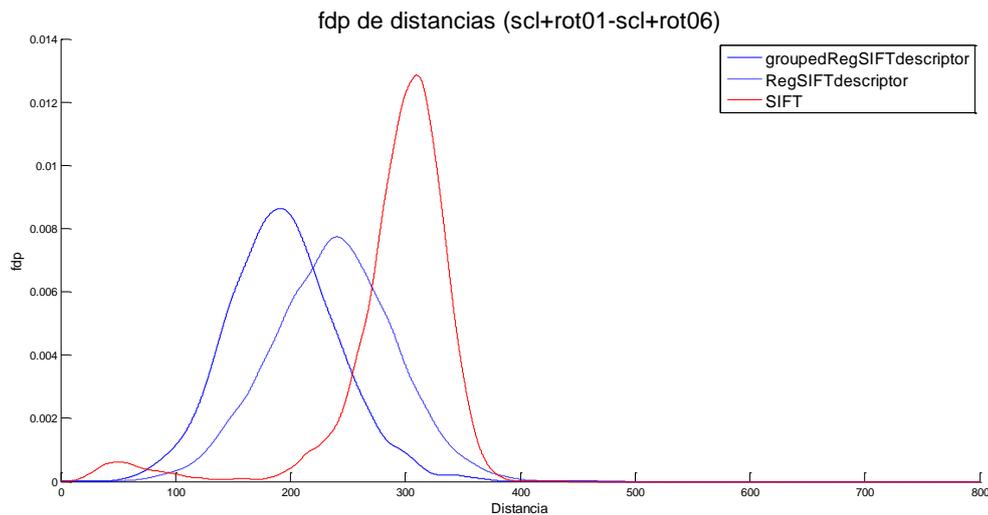


Figura 6-33: fdp de distancias para scl+rot01- scl+rot06

6.3.4. Análisis de resultados

Las conclusiones se pueden extraer a partir de las gráficas en conjunto con los resultados obtenidos en el *toy-example* sección 6.2. Para ello se ha observado lo siguiente:

- **Naturaleza de las distribuciones**

SIFT

El algoritmo de SIFT presenta una distribución de distancias cuasi-bimodal en la mayoría de las situaciones. Este comportamiento se debe principalmente a que algunos de los puntos resultan tener una descripción más precisa que otros. Las gráficas que no presentan esta distribución, y su gráfica resulta más similar a una distribución mono-modal garantizarán al menos que todos los puntos son descritos con similar robustez. En ambos casos, esto resulta positivo o negativo en función de la evolución de dichas distribuciones con los cambios en las imágenes.

RegSIFT/ groupedRegSIFT

Todas las distribuciones obtenidas para los métodos propuestos resultan presentar una distribución mono-modal. Adicionalmente, su desviación típica es, en “casi” todas las gráficas, menor que la de SIFT, por lo que se puede determinar que los puntos presentan una distribución de distancias concentrada. La consecuencia directa de este hecho es que todos los puntos están descritos con igual robustez. De nuevo, este hecho por sí mismo no contesta a la pregunta de si se mantienen o no las propiedades y en qué proporción. Se habrá de esperar a observar la evolución de la distribución con la variación de las imágenes para contestar.

- **Valores de distancias**

SIFT

El hecho de la diversidad de distribución en la descripción tiene como resultado que se presenten puntos con *matchings* a distancias muy bajas, mientras que otros por el contrario las presentan muy elevadas. Las de este segundo tipo, de incrementarse con las distintas imágenes, esto es, con el empeoramiento de las condiciones, podrían teóricamente llegar a dar lugar a error con falsos positivos presentes en esas distancias altas en aplicaciones reales.

RegSIFT/ groupedRegSIFT

Inicialmente, las distancias observadas son más elevadas que las del modo de medias bajas de la distribución de SIFT. Pese a que esto podría parecer inicialmente un empeoramiento del comportamiento, puede no resulta ser tal, pues lo importante del método no son lo valores absolutos de las distancias sino su valor relativo con posibles FP. Este valor relativo ya se comprobó que mejoraba en los resultados del *toy-example* 6.2, por lo que de nuevo habrá que considerar la evolución de las distancias para poder extraer conclusiones respecto de las propiedades.

- **La evolución de las distribuciones de distancias**

SIFT

En este punto se observa el verdadero comportamiento del método de SIFT. Como se puede observar, según las evaluaciones de las propiedades se vuelven más exhaustivas, esto quiere decir, aumenta progresivamente el *blur*, se reduce notablemente la escala, o se elimina casi toda la luminosidad, la distribución de las distancias de los *matchings* de SIFT tiende a desplazar su media notablemente hacia valores altos. Por lo tanto, pese a que en el estado del arte ya se comprobó que el comportamiento de SIFT frente a estas propiedades es superior al de otros métodos (2.3), su estabilidad no es completamente robusta.

RegSIFT/ groupedRegSIFT

A partir de las descripciones precisas, consecuencia de las distribuciones de distancias muy concentradas, la evolución de las medias resulta el punto crítico para la comprobación de la robustez del método frente a las propiedades evaluadas.

Los resultados resultan ser muy positivos, y notablemente superiores a los presentados por SIFT. Las medias de las distribuciones de distancias concentradas se mantienen muy estables a medida que se aumenta la rigurosidad de la evaluación. Esto quiere decir que las descripciones efectuadas por el método propio se ven muy poco alteradas por las variaciones estudiadas, y que por lo tanto se puede concluir que el método estudiado es robusto, incluso más que SIFT, a las variaciones de iluminación, escala, rotación y borrosidad.

En función de lo observado y del estudio realizado, se puede extraer una conclusión nítida. El método presenta una extraordinaria robustez a todas las propiedades evaluadas, superando en la mayoría de casos los comportamientos del método original.

Con la restante comprobación del comportamiento del nuevo algoritmo sobre una aplicación real, experimento que se llevará a cabo en el siguiente capítulo (Capítulo 7), y en caso de ser positiva, se podrá decir que se ha logrado desarrollar un algoritmo efectivo y preciso para la descripción de imágenes.

Capítulo 7 – Aplicaciones

7.1 Introducción

La última de las comprobaciones a las que se va a someter al algoritmo desarrollado va a ser su uso sobre aplicaciones reales. Para ello, se han desarrollado completamente desde cero dos aplicaciones, *Duplicate Image Recognition (Reconocimiento de logotipos)* y Reconocimiento ante oclusiones, similares, en objetivo, a las presentadas en [37] [47].

La elección de cada una de las aplicaciones queda plenamente justificada por los objetivos que tratan de comprobar:

- ***Duplicate Image Recognition (Reconocimiento de logotipos)***: Esta aplicación se centra en comprobar el comportamiento del algoritmo en el objetivo planteado en el Capítulo 4. Trata de valorar, en un entorno esta vez real, la capacidad del algoritmo para independizar información de las distintas regiones de la imagen.
- ***Reconocimiento ante oclusiones***: Aunque esta aplicación no se encuentra tan directamente centrada en los defectos de aplicación como la anterior, como ya se observó en la sección 4.3, si que presente cierta relación. Sin embargo, el objetivo principal del desarrollo de esta aplicación es el de comprobar el funcionamiento del método sobre una aplicación no focalizada en los objetivos perseguidos por el mismo. De esta manera se logrará comprobar el funcionamiento del algoritmo en un nivel más general.

7.2 Reconocimiento de logotipos

Las aplicaciones de reconocimiento de entidades en imágenes, o logotipos como es el caso de ejemplo analizado, están en continuo desarrollo y mejora. Su utilidad por motivos de copyright, estudios de mercados o búsquedas web ha hecho que numerosos esfuerzos de la comunidad científica se centren en ellas. Siendo pues una aplicación en desarrollo hoy en día, era de esperar que se hayan expuesto métodos que recurren al uso de puntos SIFT [37] para ello. Su comportamiento, como ya se expuso en la sección 4.2, se ve limitado por los defectos asociados a aplicación citados en dicho apartado.

7.2.1 Metodología de evaluación

7.2.2.1 Dataset

El *dataset* utilizado en esta aplicación ha sido de elaboración propia. Para su elaboración se ha recurrido a buscadores web de donde se ha extraído lo siguiente:

- 1 Set de 15 imágenes de logotipos conformado por:
 - 3 Imágenes por logotipo
 - 5 Familias de logotipos

Las distintas familias de logotipos incluidos en el *dataset* han sido las mostradas a continuación. El resto de imágenes pueden verse en el Anexo A.C

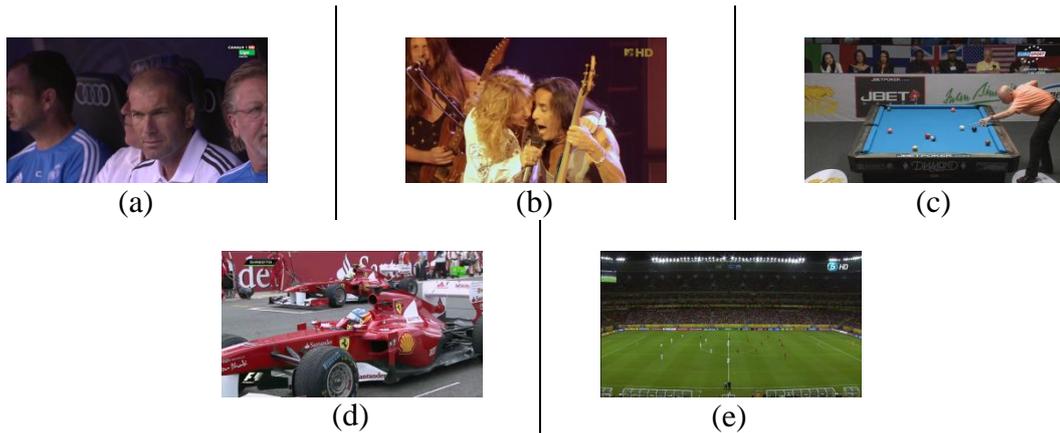


Figura 7-1: Imágenes por familias de logotipos. a)Canal+HD b)MTV-HD c)EuroSport d)LaSexta HD e)Telecinco HD Fuente: Propia⁷

7.2.2.2 Métrica

A nivel de métrica es necesario determinar los métodos seguidos en dos aspectos:

- Cálculo de distancias y establecimiento de *matchings* descriptores.
- Obtención de los scores de cada imagen con las distintas familias de *train*.

Matchings

Los cálculos de distancias y *matchings* se han realizado según lo indicado en la sección 5.5.3, Ecuación 13 Ecuación 14 y Ecuación 15.

Scores

Para la extracción de resultados, se puntuarán como positivas las 30 *features* de distancias menores. Este número se mantendrá fijo para todas las pruebas.

7.2.2 Diseño

La aplicación desarrollada sigue un esquema de funcionamiento dividido en etapas similar al mostrado en la Figura 7-2



⁷ Todas las imágenes expuestas son propiedad de marcas registradas, y su uso es únicamente con motivos científicos. Su localización se realizó a través del buscador web de imágenes *Google*.

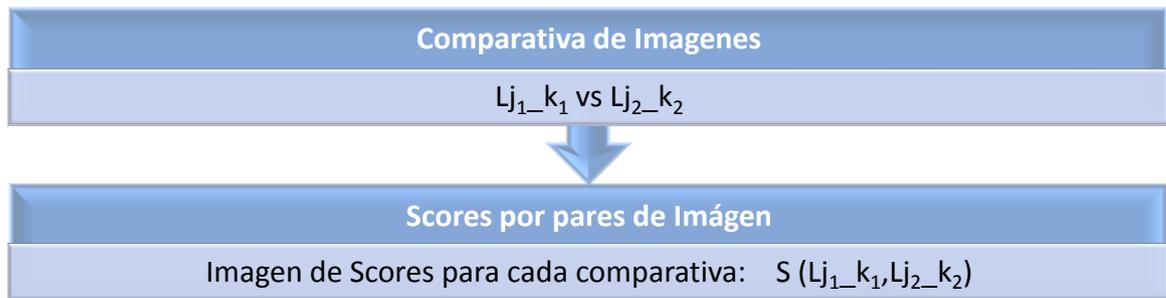


Figura 7-2: Esquema de diseño de la aplicación. Fuente: Propia

Etapa por etapa, el funcionamiento de la aplicación será:

a) *Imágenes de entrada:*

Ser recibirán las 15 imágenes detalladas en el *dataset*, sección 7.2.2.1, y se clasificarán como

$$L_{j-k}$$

j la familia del logo

k la imagen de la familia

b) *Descripción RegSIFT:*

Cada imagen recibida será procesada por el algoritmo desarrollado obteniéndose los descriptores *RegSIFT* correspondientes a cada imagen.

c) *Matching de Imágenes:*

Se procederá a realizar el *matching* de los puntos según lo indicado en la sección 7.2.2.2, entre pares de imágenes $L_{j_1-k_1}$ y $L_{j_2-k_2} \forall \{j_1-k_1\} \neq \{j_2-k_2\}$, esto es, entre todas las imágenes pero evitando la comparación de una imagen consigo misma.

d) *Scores*

Los resultados serán imágenes de *Scores* $S(L_{j_1-k_1}, L_{j_2-k_2})$ resultantes de la activación de las regiones asociados a los 30 *matching* de menor distancia para cada par de imágenes entre las que se realizó una comparativa. Para cada uno de estos pares, se obtendrá pues una imagen de *Scores* con las citadas regiones activadas y el resto desactivadas.

7.2.3 Resultados

Como resultados de la aplicación, y tal y como se indico en la sección de diseño 7.2.2 apartado d), serán unas imágenes de *Scores* para cada par de imágenes confrontadas. En dichas imágenes se podrá ver el resultado de las regiones activadas superpuesto sobre la imagen con la que se ha llevado a cabo la confrontación o *matching*.

Con el fin de ser exhaustivos pero no redundantes, se mostrará la comparativa de un representante de cada familia frente al resto de las imágenes del *dataset*.

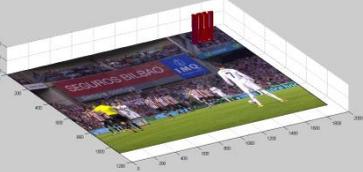
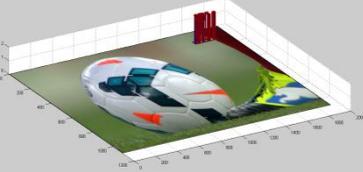
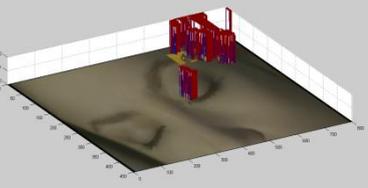
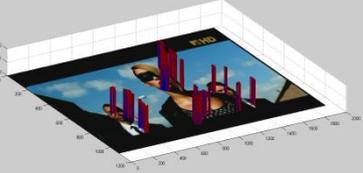
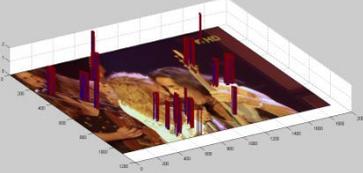
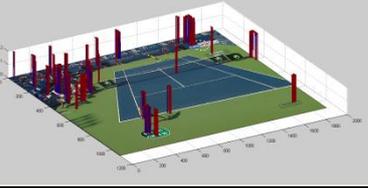
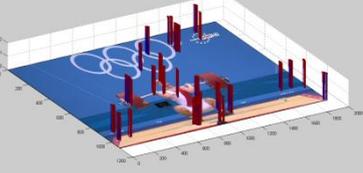
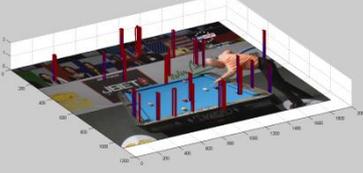
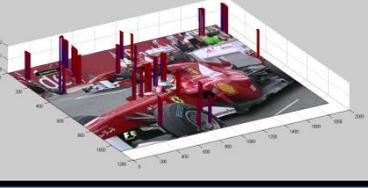
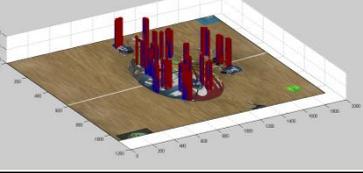
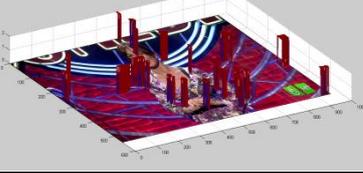
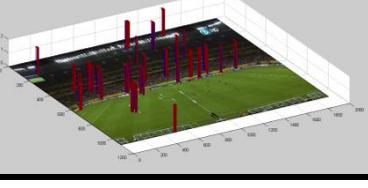
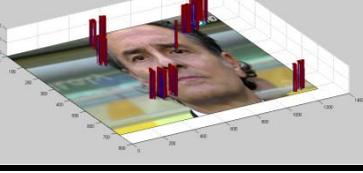
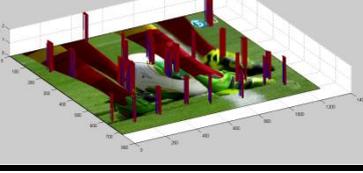
Representante L1_1	S(L1_1,L1_2)	S(L1_1,L1_3)
		
S(L1_1,L2_1)	S(L1_1,L2_2)	S(L1_1,L2_3)
		
S(L1_1,L3_1)	S(L1_1,L3_2)	S(L1_1,L3_3)
		
S(L1_1,L4_1)	S(L1_1,L4_2)	S(L1_1,L4_3)
		
S(L1_1,L5_1)	S(L1_1,L5_3)	S(L1_1,L5_3)
		

Tabla 7-1: Resultados para representante de *familia1*. Fuente: Propia

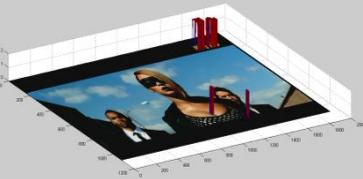
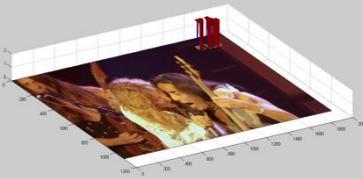
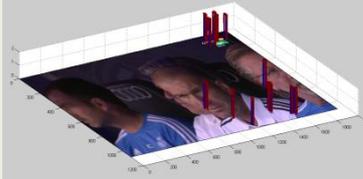
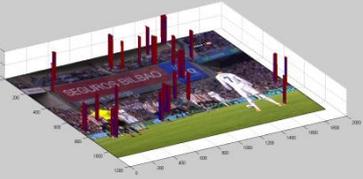
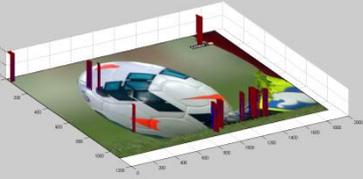
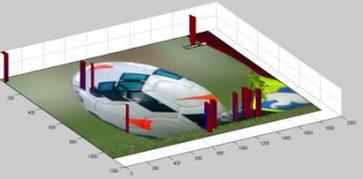
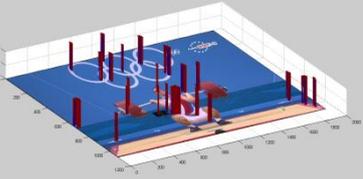
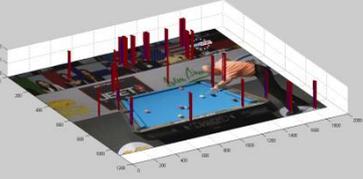
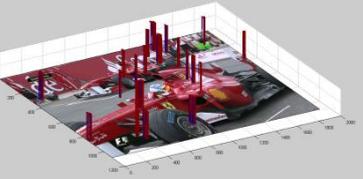
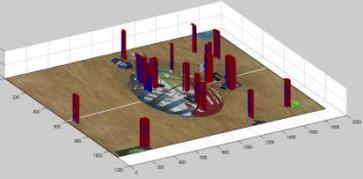
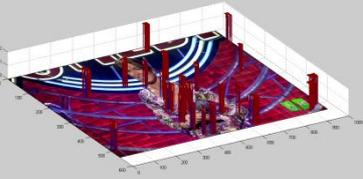
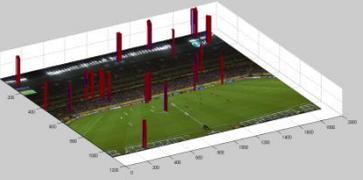
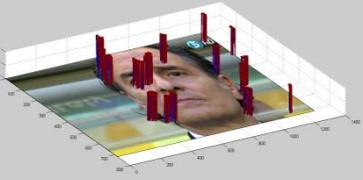
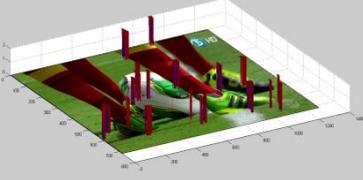
Representante L2_1	S(L2_1,L2_2)	S(L2_1,L2_2)
		
S(L2_1,L1_1)	S(L2_1,L1_2)	S(L2_1,L1_3)
		
S(L2_1,L3_1)	S(L2_1,L3_2)	S(L2_1,L3_3)
		
S(L2_1,L4_1)	S(L2_1,L4_2)	S(L2_1,L4_3)
		
S(L2_1,L5_1)	S(L2_1,L5_2)	S(L2_1,L5_3)
		

Tabla 7-2: Resultados para representante de familia2. Fuente: Propia

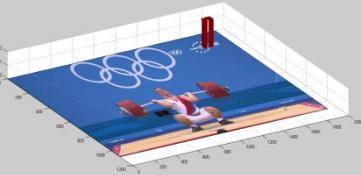
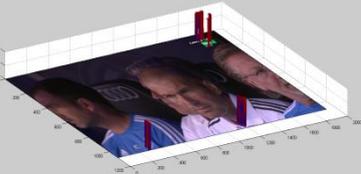
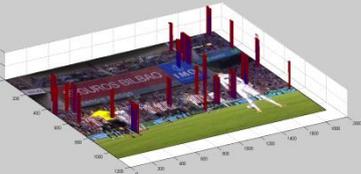
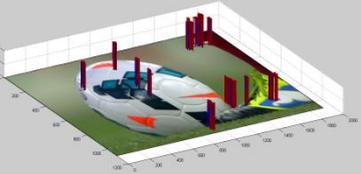
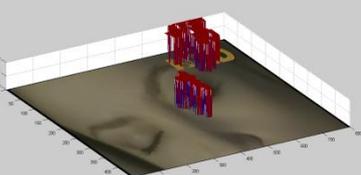
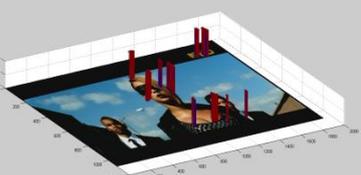
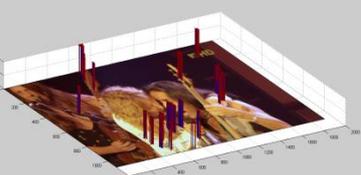
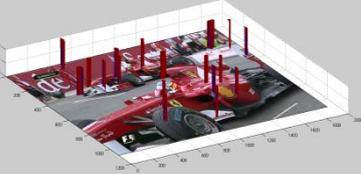
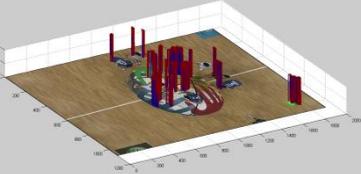
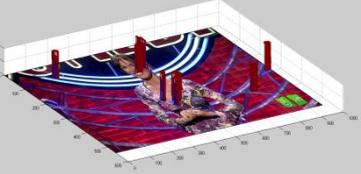
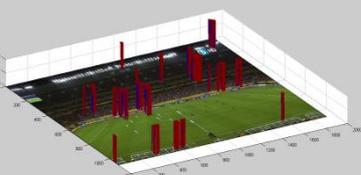
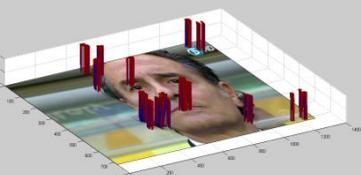
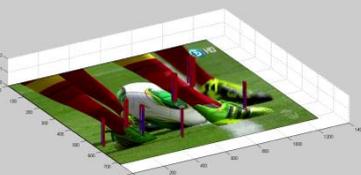
Representante L3_1	S(L3_1,L3_2)	S(L3_1,L3_3)
		
S(L3_1,L1_1)	S(L3_1,L1_2)	S(L3_1,L1_3)
		
S(L3_1,L2_1)	S(L3_1,L2_2)	S(L3_1,L2_3)
		
S(L3_1,L4_1)	S(L3_1,L4_2)	S(L3_1,L4_3)
		
S(L3_1,L5_1)	S(L3_1,L5_2)	S(L3_1,L5_3)
		

Tabla 7-3: Resultados para representante de familia3. Fuente: Propia

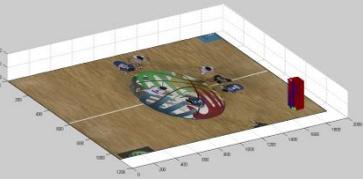
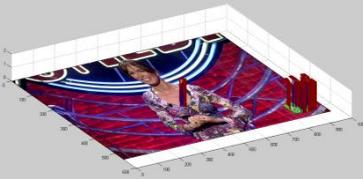
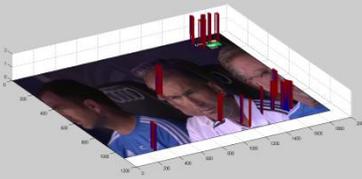
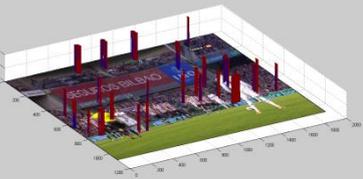
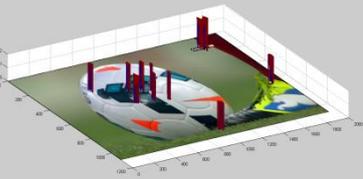
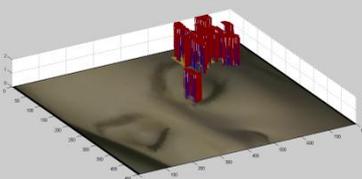
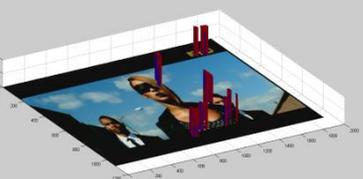
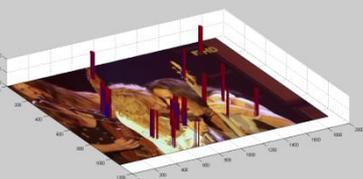
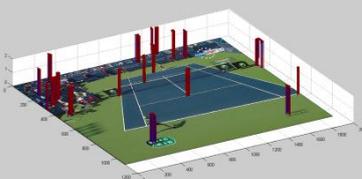
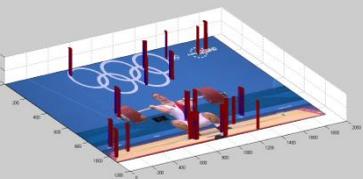
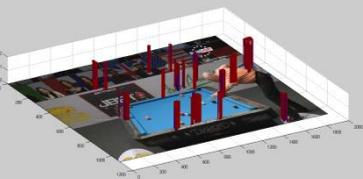
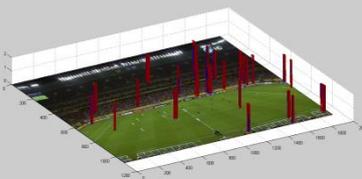
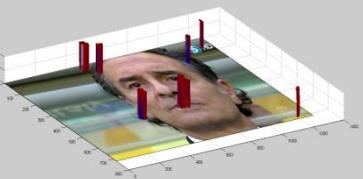
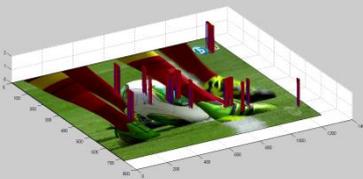
Representante L4_1	S(L4_1,L4_2)	S(L4_1,L4_3)
		
S(L4_1,L1_1)	S(L4_1,L1_2)	S(L4_1,L1_3)
		
S(L4_1,L2_1)	S(L4_1,L2_2)	S(L4_1,L2_3)
		
S(L4_1,L3_1)	S(L4_1,L3_2)	S(L4_1,L3_3)
		
S(L4_1,L5_1)	S(L4_1,L5_2)	S(L4_1,L5_3)
		

Tabla 7-4: Resultados para representante de familia4. Fuente: Propia



Tabla 7-5: Resultados para representante de familia5. Fuente: Propia

7.2.4 Análisis de resultados y evaluación

Con el fin de extraer unas conclusiones del funcionamiento de la aplicación fiables, se va a llevar un análisis de los resultados obtenidos en la sección 7.2.3 en base a dos criterios: Intrasimilitud (capacidad para localizar su propio logo) e Interseparabilidad (capacidad para no confundirse con otros).

Resultados Intrasimilitud (L_{j_1-1} vs L_{j_1-k})

En la gran mayoría de casos, se pueden observar como las regiones activadas corresponden con el logo de la familia. Esto se debe a que el método aísla la

información de las regiones de los logos, y las localiza casi inequívocamente sobre las imágenes de la familia en las que vuelve a aparecer el mismo logo.

La precisión no es del 100 % como se observan en los errores puntuales cometidos por ejemplo en la Tabla 7-2 imagen $S(L2_1, L2_2)$ o en la Tabla 7-5 imagen $S(L5_1, L5_3)$.

Resultados Interseparabilidad (L_{j1-1} vs L_{j2-k})

En estos casos se observan dos comportamientos diferenciados. Por una parte, en las imágenes con alta información las regiones activadas están dispersas, puesto que los *matchings* existentes no resultan representativos. Sin embargo, en imágenes poco texturadas o de baja información, como es el caso de la Figura 0-39, los *matchings* tienden a concentrarse en las zonas de mayor información como el propio logotipo (ver Tabla 7-1 imagen $S(L1_1, L2_1)$).

En base al objetivo de la aplicación, y a partir de los resultados y su análisis, se pueden extraer las siguientes conclusiones sobre la misma.

- El reconocimiento de logotipos en televisión no resulta una tarea trivial, afrontándose principalmente de 2 maneras, todas ellas fuertemente limitadas:
 - Correlaciones de Bloque: útiles para cuando las imágenes tienen el mismo tamaño, formato, resolución y orientación.
 - *Matching* de descriptores: la mayoría de ellos, como SIFT, incorporan información del entorno, por lo que si el logo aparece sobre distintos fondos suele realizarse una descripción errónea, especialmente si los fondos contienen mucha información.
- La estrategia seguida por la aplicación y el algoritmo corresponde con el segundo punto, lográndose invariencias a cambios de escala y resolución. Adicionalmente, las propiedades del algoritmo consiguen independizar informaciones de distintas regiones, por lo que la descripción de los logos no se verá afectada por los fondos sobre los que aparecen.
- En las imágenes de resultados se pueden observar claras evidencias de este comportamiento. La reconstrucción de los logotipos en la mayoría de las ocasiones es casi completa y no incorpora información alguna de los fondos. Los resultados para las imágenes en las que los logotipos presentan variaciones de escala o resolución presentan iguales resultados.

A partir de estas ideas, como conclusión global de la aplicación se puede decir que, en un ámbito más abierto y real, el algoritmo ha presentado las propiedades y resultados esperados: alta distintibilidad, alta repetitibilidad, propiedades de invariencias similares a las de SIFT y capacidad de describir independientemente regiones no relacionadas.

Como contra, cabe destacar que las imágenes utilizadas son de alta calidad y resolución, pues el método desarrollado resulta dependiente de la segmentación, la cual funciona de forma más estable cuando las imágenes no se encuentran degradadas, e.g. por compresión o codificación.

7.3 Reconocimiento ante oclusiones

El reconocimiento de elementos de una imagen, y más concretamente su detección e identificación sobre distintas imágenes o frames de un vídeo, son la base de numerosas tareas de alto nivel como el seguimiento de objetos y personas, el reconocimiento de movimientos o el análisis de anomalías en secuencias de vídeo. Sin embargo, en todas ellas, la detección de un mismo elemento sobre distintas imágenes resulta ser una tarea nada trivial. Entre los principales problemas a los que se enfrenta destacan las oclusiones.

Definiéndose como oclusión la ocultación de todo o parte del elemento buscado entre dos imágenes, se estudiará el comportamiento del algoritmo desarrollado y del de SIFT en una aplicación de evaluación de imágenes con elementos ocluidos.

7.3.1 Metodología de evaluación

7.3.2.1 Dataset

El *dataset* empleado ha sido desarrollado explícitamente para este proyecto. Para ello se han seguido los siguientes patrones:

- Total de 19 imágenes de alta resolución tomadas en tres categorías con sus correspondientes *ground-truth hand-annotated*, a saber:
 - 1 Imagen referencia del objeto.
 - 6 Imágenes de objeto ocluido con elemento poco texturado (categoría OclusPT).
 - 6 Imágenes de objeto ocluido con elemento altamente texturado (categoría OclusAT).
 - 6 Imágenes de evaluación de oclusiones de dos tipos y cambios de propiedades (categoría OclusPDo).

A continuación se muestra unos ejemplos de cada una de las tres categorías:

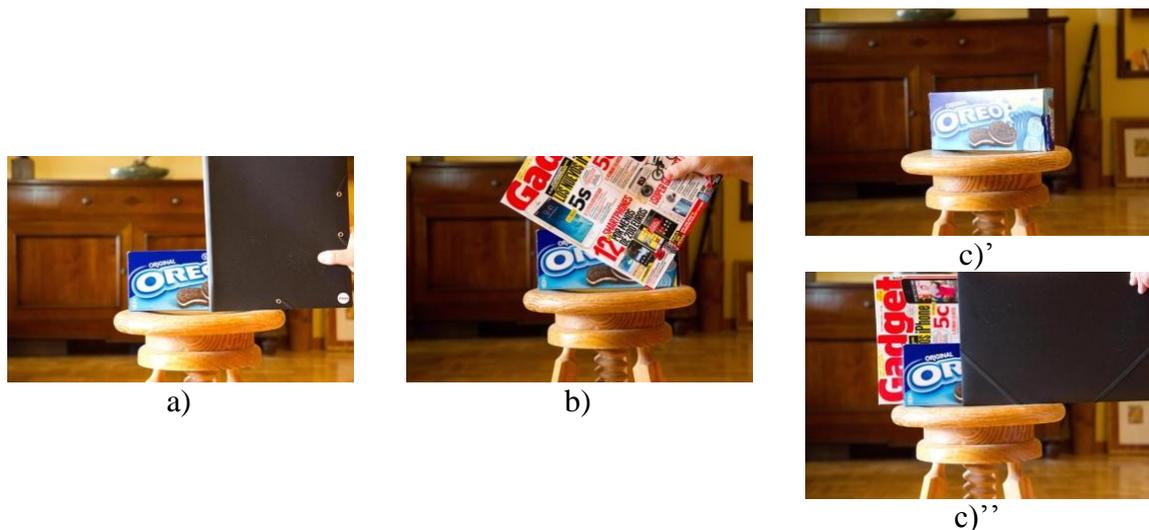


Figura 7-3: Imágenes por categorías. a) Ocluido Poco Texturado b) Ocluido Altamente Texturado c)' Evaluación de propiedades c)'' Ocluido Dos elemento Fuente: Propia

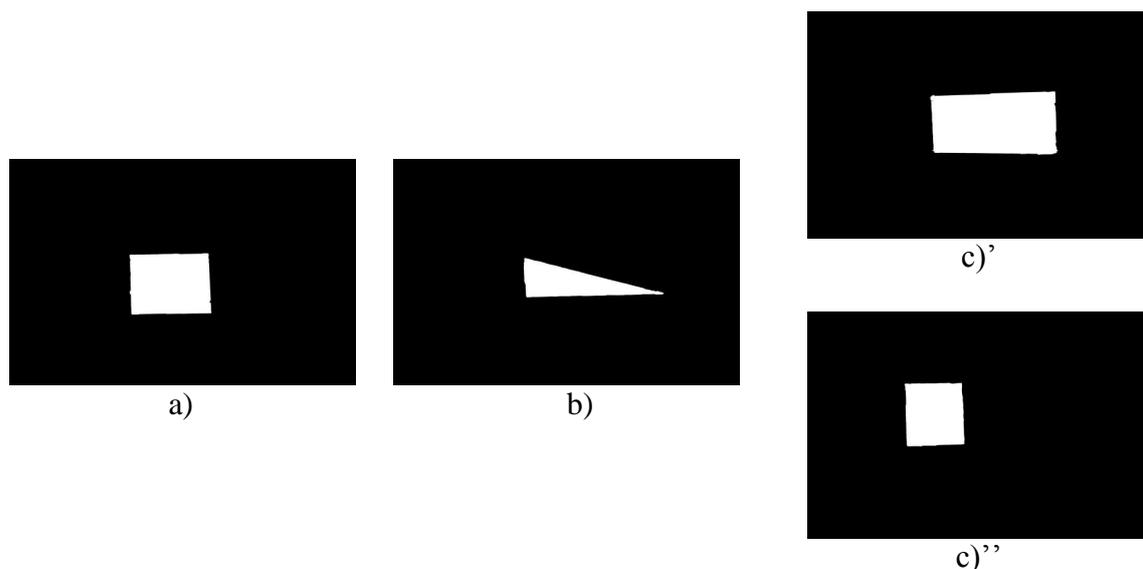


Figura 7-4: *Ground-truth* por categorías. a) Ocluido Poco Texturado b) Ocluido Altamente Texturado c)' Evaluación de propiedades c)'' Ocluido Dos elemento Fuente: Propia

Todas las imágenes incluidas en el *dataset* de la aplicación se encuentran en la sección correspondiente del Anexo A.A

7.3.2.2 Métrica

A nivel de métrica es necesario determinar los métodos seguidos en dos aspectos:

- Cálculo de distancias y establecimiento de *matchings* de descriptores.
- Obtención de los resultados cuantitativos de cada secuencia.

Matchings

Los cálculos de distancias y *matchings* se han realizado de distintas maneras según la categoría de análisis.

Para la evaluación de las categorías de oclusiones con Elementos Poco Texturados, Altamente Texturados y de Propiedades y Dobles Oclusiones, los *matchings* se han realizado punto a conjunto según lo indicado en la sección 5.5.2, Ecuación 13, Ecuación 14 y Ecuación 15, entre los descriptores de la imagen de referencia y los descriptores de las imágenes de la categoría evaluada.

Scores

Para la extracción de datos se considerarán las localizaciones, frente o fondo, de los distintos puntos involucrados en cada par de *matching*. En función de las distintas combinaciones posibles se realizarán las siguientes clasificaciones:

TruePositives: Las *features* que han originado el *matching* se encuentran ambas localizadas sobre el objeto de interés.

FalsePositives: La *feature* de la imagen de referencia se encuentra localizada fuera del objeto de interés, en el fondo, y su par de *matching* por el contrario se ubica en el objeto de la testada.

FalseNegative: La *feature* de la imagen de referencia se encuentra en el objeto de interés, mientras que su par correspondiente se localiza en el fondo.

Con la intención de llevar a cabo un análisis en las regiones de más interés, se generará una clase borde, en la que se definirán las mismas categorías de datos que las citadas en el párrafo anterior, pero donde las *features* de la imagen de test que se evaluarán serán únicamente las ubicadas en el borde creado entre el objeto de frente y el elemento oclusivo. Así, aparecerán las categorías: **TruePositivesBorde, FalsePositiveBorde y FalseNegativeBorde.**

7.3.2 Diseño

La aplicación en si misma tiene como objetivo la identificación de un objeto en distintas imágenes en las cuales presenta distintos eventos y niveles de oclusión. Para ello, se ha diseñado un sistema básico con las funcionalidades mostradas en la Figura 7-5.

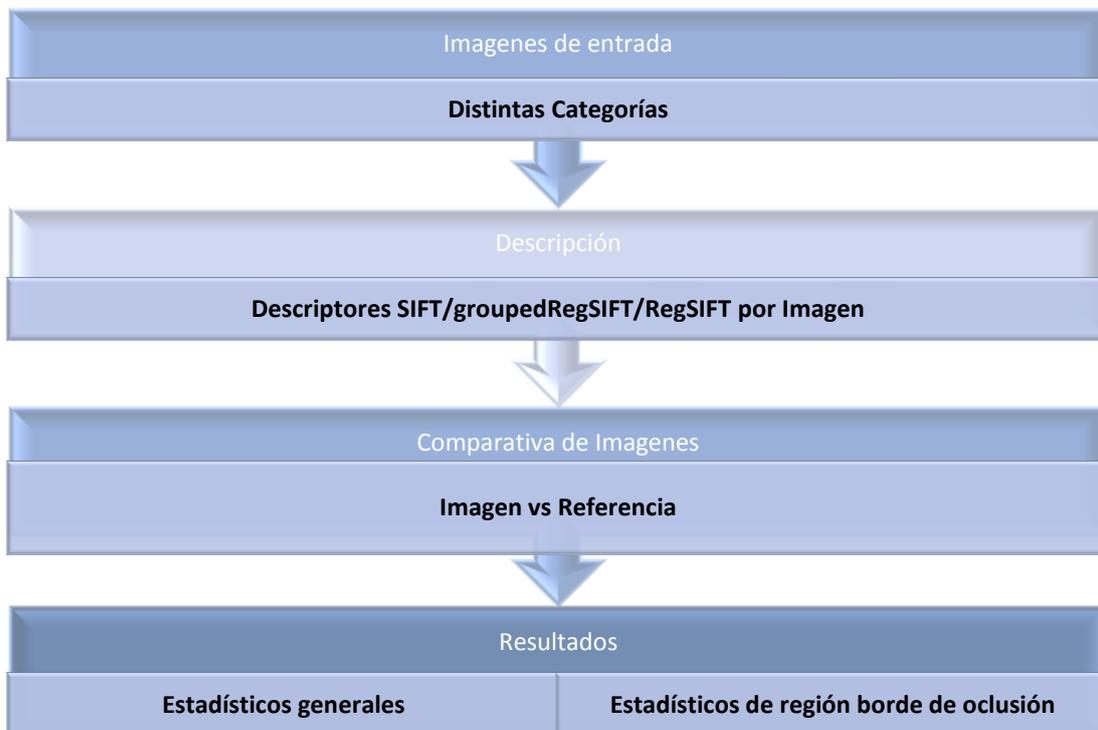


Figura 7-5: Esquema de etapas de aplicación de oclusiones. Fuente: Propia

Se comienza por una etapa previa en la que se recibe y describe la imagen de referencia del objeto a estudiar sin oclusión alguna.

A continuación, etapa por etapa, el funcionamiento de la aplicación será:

a) *Imágenes de entrada:*

Se recibe, tal y como se indicó al pie de la figura, una imagen de referencia del objeto y 18 imágenes pertenecientes a tres categorías distintas tal y como fueron descritas en la sección 7.3.2.1.

b) *Descripción:*

Cada imagen recibida será procesada por el algoritmo original, SIFT, y por el desarrollado, obteniéndose los descriptores SIFT, *groupedRegSIFT* y *RegSIFT* sección 5.3.4.

c) *Comparativa de Imágenes:*

Siempre respecto a la de referencia, se procederá a realizar el *matching* de los puntos según lo indicado en la sección 7.3.2.2. Como resultado de este *matching* se asignará a cada par de puntos asociados de las imágenes comparadas una distancia.

d) *Scores*

Los resultados, tal y como se ha anticipado en la Figura 7-5, serán de dos tipos, a saber:

Estadísticos generales: En esta sección se evaluará el sistema al completo. Se extraerán valores de TruePositives (TP), FalsePositives (FP) y FalseNegatives(FN) de toda la imagen, a partir de los cuales se extraerán los resultados generales de la aplicación, sección 7.3.1.1.

Estadísticos de región borde de oclusión: Al igual que la sección de estadísticos generales, se extraerán valores de TP, FP y FN. Sin embargo, para estos datos solo se tendrán en cuenta aquellos que involucren *features* que se localicen en las zonas del borde del elemento oclusivo sobre el objeto de frente.

7.3.3 Resultados

Los datos extraídos, por su distinta naturaleza de su dan lugar a dos tipos de resultados que se estudiarán a continuación:

- **Resultados generales:** Se observarán gráficamente los valores de Precisión, *Recall* y *FScore* para las distintas categorías de entrada descritas con ambos métodos.
- **Resultados de región borde de oclusión:** Resultados similares a los generales, pero orientados únicamente a observar la influencia de los nuevos bordes e informaciones introducidos por las oclusiones.

7.3.1.1 Resultados generales

Los resultados obtenidos para los tres objetivos de estudio, Precisión, *Recall* y *FScore* han sido los siguientes:

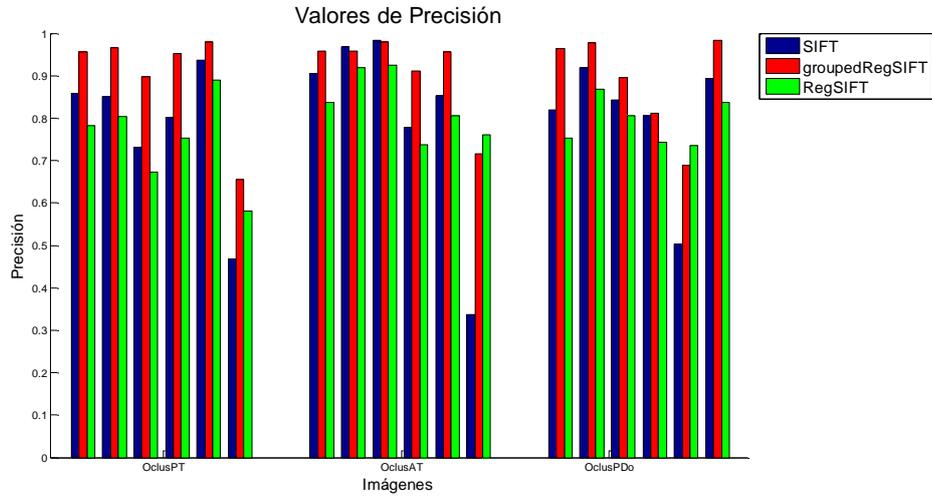


Figura 7-6: Valores de precisión generales por categoría y método. Fuente: Propia

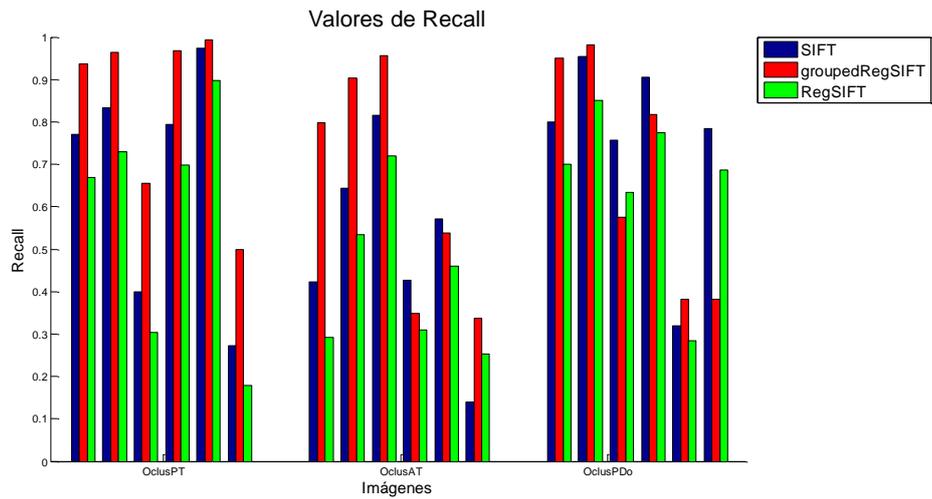


Figura 7-7: Valores de *Recall* generales por categoría y método. Fuente: Propia

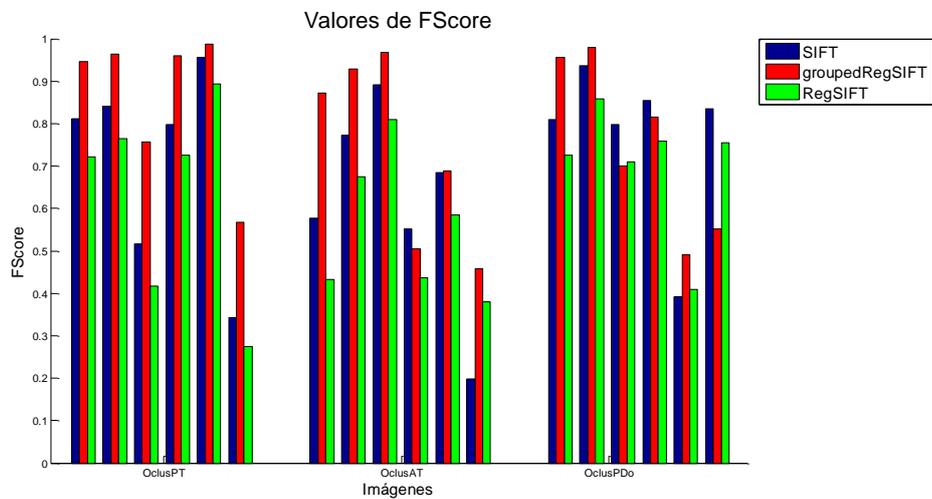


Figura 7-8: Valores de *FScore* generales por categoría y método. Fuente: Propia

Con la intención de observar de un modo más general el comportamiento del sistema se incluye la siguiente tabla de valores medios por categoría a partir de lo mostrado en las gráficas expuestas anteriormente.

	Precisión <small>media</small>			Recall <small>medio</small>			FScore <small>medio</small>		
	OclusPT	OclusAT	OclusPDoT	OclusPT	OclusAT	OclusPDoT	OclusPT	OclusAT	OclusPDoT
<i>SIFT</i>	0.7740	0.8039	0.7808	0.6741	0.5034	0.7534	0.7113	0.6125	0.7704
<i>Grouped-RegSIFT</i>	0.9071	0.9130	0.8872	0.8358	0.6468	0.6810	0.8638	0.7366	0.7487
<i>RegSIFT</i>	0.7472	0.8308	0.7903	0.5797	0.4277	0.6549	0.6329	0.5529	0.7027

Tabla 7-6: Valores medios de Precisión, Recall y FScore generales aplicación de Oclusiones.
Fuente: Propia

Adicionalmente, se ha considerado positivo incluir el porcentaje de mejora del sistema respecto del comportamiento de SIFT en términos del *FScore* introducido por el método con el descriptor *groupedRegSIFT*. Los valores obtenidos de esta medición han sido:

	%Mejora <i>FScore</i>		
	OclusPT	OclusAT	OclusPDoT
<i>Grouped-RegSIFT</i>	21,44	20,26	-2.81^(*)

Tabla 7-7: Porcentajes de mejora FScore en aplicación de Oclusiones Fuente: Propia

(*) Este valor negativo se debe a una imagen con *Recall* muy por debajo de la media. Abstrayendo este fallo puntual, pues la imagen Figura 0-79 no es representativa ni un caso único, el porcentaje de mejora subiría a 5,45% en la secuencia que evalúa dos imágenes de dobles oclusiones y cuatro de variaciones de condiciones de la imagen sin oclusiones.

7.3.1.2 Resultados de región borde de oclusión

Una vez expuestos los resultados generales de la aplicación se va a presentar a continuación los resultados obtenidos para el estudio realizado en las zonas de interés.



Figura 7-9: Detalle de región borde de oclusión. Fuente: Propia

Gráficamente, los valores de Precisión, *Recall* y *FScore* resultantes han sido⁸:

⁸ Las imágenes 1-4 de la categoría OclusPDo no presentan oclusiones y por lo tanto tampoco puntos de borde de oclusión, y por ende no serán analizadas.

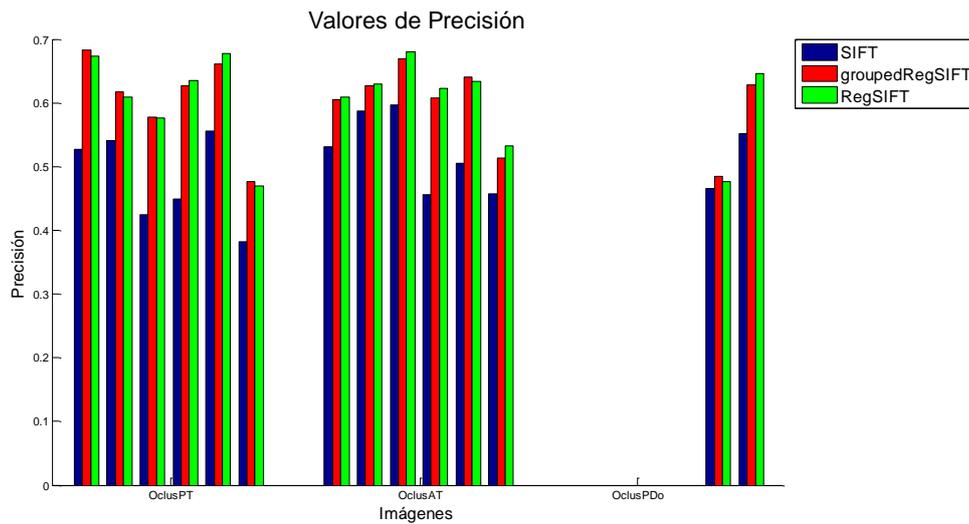


Figura 7-10: Valores de precisión en borde de oclusión por categoría y método. Fuente: Propia

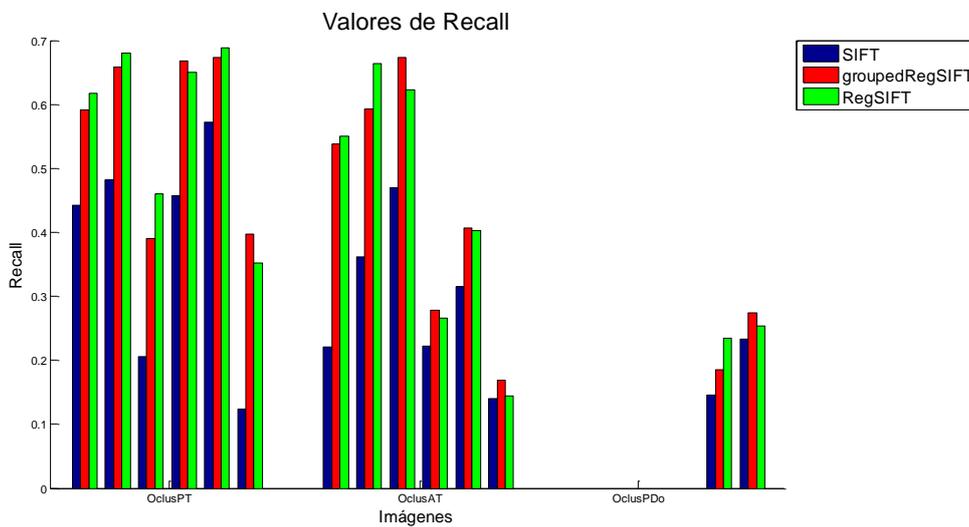


Figura 7-11: Valores de Recall en borde de oclusión por categoría y método. Fuente: Propia

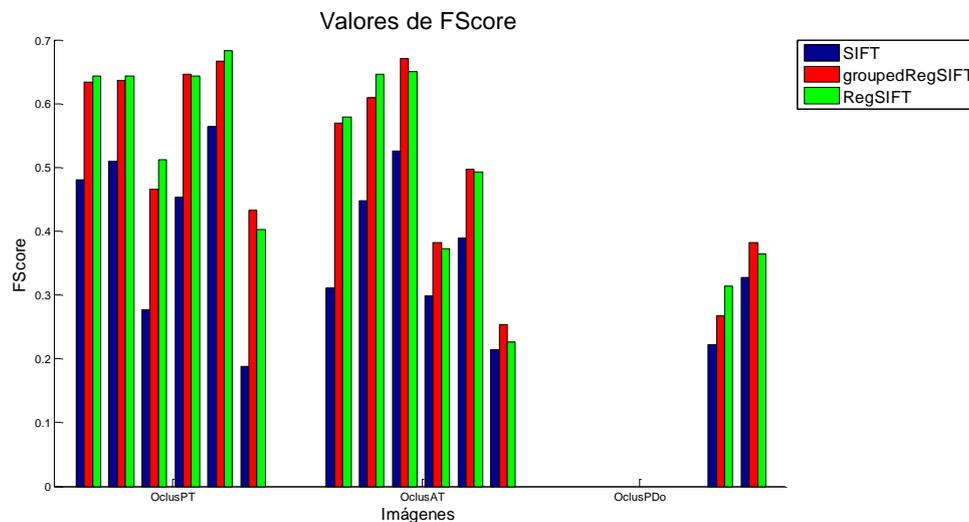


Figura 7-12: Valores de FScore en borde de oclusión por categoría y método. Fuente: Propia

Con la intención de observar el comportamiento general nuevamente se adjunta una tabla con los valores medios por secuencia de los resultados mostrados en las gráficas.

	Precisión <small>media</small>			Recall <small>medio</small>			FScore <small>medio</small>		
	OclusPT	OclusAT	OclusPDoT	OclusPT	OclusAT	OclusPDoT	OclusPT	OclusAT	OclusPDoT
<i>SIFT</i>	0.4800	0.5222	0.5126	0.3805	0.2884	0.3699	0.4121	0.3645	0.2747
<i>Grouped-RegSIFT</i>	0.6069	0.6107	0.5976	0.5630	0.4429	0.4658	0.5806	0.4973	0.3248
<i>RegSIFT</i>	0.6070	0.6181	0.5949	0.5743	0.4415	0.4726	0.5876	0.4944	0.3390

Tabla 7-8: Valores medios de Precisión, Recall y FScore en borde de oclusión, aplicación de Oclusiones. Fuente: Propia

De igual manera que en el apartado anterior, sección 7.3.1.1, se van a incluir los porcentajes de mejora respecto del *FScore*, pero en esta ocasión de ambos tipos de descriptores de *RegSIFT*.

	%Mejora <i>FScore</i>		
	OclusPT	OclusAT	OclusPDoT
<i>Grouped-RegSIFT</i>	40.93	36.38	18.23
<i>RegSIFT</i>	42.58	35.63	23.40

Tabla 7-9: Porcentajes de mejora FScore en aplicación de Oclusiones Fuente: Propia

7.3.4 Análisis de resultados

Previa afirmación alguna sobre el funcionamiento del algoritmo en la aplicación desarrollada, se va a proceder a realizar un análisis de los resultados obtenidos en la sección 7.3.3.

Resultados generales

Evaluando la precisión, el algoritmo de SIFT presenta unos valores cercanos al 80%, lo cual son unos valores muy elevados en categorías con oclusiones. Sin embargo, la combinación de descriptores *RegSIFT* en *groupedRegSIFT* mejora las puntuaciones obtenidas por este, pues combina la robustez en bordes de *RegSIFT* con una mayor precisión en el frente que sus componentes individuales.

Analizando los datos de *Recall*, se observa como cuando el objeto oclusivo presenta más texturas origina mayor número de FP. El *Recall* en ese caso baja en todos los métodos. Sin embargo, salvo por el caso puntual, citado anteriormente, donde el *Recall* empeora para una imagen, las medias obtenidas para el método propio con el descriptor *groupedRegSIFT* mejoran, aunque en menor medida que con la precisión, los comportamientos de SIFT.

Por lo tanto, los resultados obtenidos en los *FScore* y las medidas de las mejoras porcentuales muestran que el método propio, gracias a la robustez en los bordes y a la mejora del descriptor más preciso y con mejor *Recall* en el frente, presenta mejores resultados que los obtenidos por el método de SIFT en este tipo de aplicaciones.

Resultados de región borde de oclusión

Con la intención de resaltar el correcto funcionamiento del método desarrollado en las zonas de especial interés, como son los bordes del frente, se llevó a cabo este experimento de cuyos resultados se puede exponer que:

- Los puntos SIFT pierden robustez notablemente. Se observa de una forma más acentuada en el *Recall*, aunque la Precisión también se ve afectada. Esto corrobora lo expuesto en el Capítulo 4 sobre los defectos asociados a aplicación de SIFT.
- El método propio por el contrario se mantiene estable en estas regiones, traducándose estos efectos en unas notables mejoras porcentuales de los *FScore*. Cabe destacar, que el descriptor en estas zonas no resulta tan influyente, aspecto del cual se pueden extraer una conclusión fundamental:

La caracterización de las *features* de borde del modo propuesto resulta en *features* extremadamente estables y robustas, que permiten obtener buenos resultados de manera casi independiente al descriptor. La mejora de este servirá, como se ha observado en el caso general, para obtener comportamientos también positivos en el frente.

Tomando como base lo expuesto en las líneas anteriores, y teniendo en cuenta el diseño llevado a cabo para la aplicación, se pueden extraer algunas conclusiones que se presentan a continuación:

- Las oclusiones dificultan enormemente las tareas de reconocimiento, si bien es cierto, como se ha citado en el punto anterior, que al método de SIFT le afectarán especialmente. Por el contrario, la capacidad de aislamiento en regiones conseguida en el algoritmo de *RegSIFT* permitirá que al menos parte de los puntos del borde de la oclusión sigan encontrándose, por lo que los resultados obtenidos serán superiores.
- El comportamiento del método desarrollado en las zonas de interés resulta muy positivo y superior al presentado por SIFT tal y como se pretendía lograr.
- El comportamiento observado y analizado en el general de la aplicación permite extraer dos conclusiones claras. Frente al método de SIFT, el hecho de que el número total de puntos de borde de oclusión respecto del total sea bajo hace que la mejora percibida no resulte tan notable. Entre los tipos de descriptores, se puede concluir que el descriptor *groupedRegSIFT* complementa muy positivamente las propiedades del algoritmo implementado en este tipo de aplicaciones.

En definitiva, se ha podido observar con esta aplicación que, para situaciones en las que se produzcan oclusiones, ciertamente bastante generalizadas, el método presentado obtiene unos resultados superiores a los presentados por SIFT y por tanto, se ha logrado paliar en cierta medida el defecto de aplicación expuesto en la sección 4.3.

Capítulo 8 – Conclusiones y trabajo futuro

8.1 Conclusiones

Este proyecto afrontaba el reto de desarrollar un algoritmo que tenía como objetivo mejorar el comportamiento del mejor de los métodos de extracción y descripción de puntos de interés del estado del arte del tratamiento de imágenes.

Para este fin, se llevó a cabo un estudio de los métodos del estado del arte, analizando los mejores detectores y descriptores según los artículos comparativos de referencia en el campo.

Tras este estudio, se ha extraído la conclusión de que los puntos SIFT son, por diversos motivos que van desde el número de referencias hasta el comportamiento frente a variaciones de las propiedades de una imagen, los mejores para la descripción de imágenes.

Puesto que el objetivo del proyecto era mejorar el comportamiento de este método, se llevó a cabo un análisis muy detallado del funcionamiento del mismo.

Tras este análisis, fue posible llevar a cabo un análisis del comportamiento del mismo en diversas aplicaciones con el fin de detectar defectos asociados a dicha aplicación que pudiesen ser corregidos.

En esta línea, se concluye que los puntos SIFT presentan defectos asociados a aplicación en su descripción por el hecho de incluir en ella información del entorno. Todas las aplicaciones que describan elementos en entornos cambiantes se verán pues afectadas por estos defectos.

Localizado el objetivo a cubrir con el algoritmo, se decidió incorporar el uso de segmentadores, concretamente el de *superpíxeles*, sobre la descripción de SIFT con el fin de poder lograr un algoritmo que describa las distintas regiones que confluyen en un mismo punto de una imagen de forma independiente.

Para comprobar que el algoritmo propuesto cumplía con el objetivo buscado y mantenía las propiedades y bondades de su predecesor, se llevó a cabo un *toy-example* y un estudio de propiedades siguiendo el guión de los citados artículos comparativos de referencia en el campo.

Tras este punto, se puede extraer la conclusión de que, en entornos preparados para ello, el algoritmo implementado logra llevar a cabo el objetivo para el que fue diseñado, y mantiene estables e incluso mejora, las propiedades positivas ya presentes en el algoritmo de SIFT.

Sin embargo, aun era necesario comprobar que el comportamiento presentado tan positivamente se mantenía para entornos reales. Con este objetivo, se diseñaron dos aplicaciones sobre las que incorporar el algoritmo desarrollado y poder evaluar su comportamiento.

En primer lugar, de la aplicación de reconocimiento de logotipos, *Duplicate Image Recognition*, se han extraído comportamientos cualitativos, puesto que no existen en la

actualidad métodos que implementen funcionamientos similares al desarrollado con el que comparar los resultados. Sin embargo, a partir de los resultados y sus análisis se ha podido sacar una serie de conclusiones principales, a saber:

- Ha resultado de gran interés la incorporación de regiones a la descripción de SIFT. No únicamente desde el punto de vista del funcionamiento básico de SIFT, sino también porque posibilita extraer resultados en función de regiones como los mostrados en la sección de resultados de la aplicación.
- En aplicaciones en las que los defectos de aplicación enunciados son críticas para el funcionamiento se ha podido observar como el funcionamiento del algoritmo permite obtener excelentes resultados cualitativos.
- Las condiciones de captura, calidad de la imagen y resoluciones de las mismas resultan factores críticos para el correcto funcionamiento del segmentador introducido, y en definitiva del algoritmo.

Por otra parte, la aplicación de reconocimiento de objetos con oclusión se ha evaluado bajo el algoritmo de SIFT y el propio. El objetivo de la misma era comprobar que el algoritmo, por el hecho de incorporar las propiedades específicas para la corrección de los citados defectos de aplicación no había perdido las restantes capacidades de SIFT de trabajar de manera robusta en aplicaciones más genéricas.

- A la vista de los resultados obtenidos, se ha demostrado que el comportamiento general se ha mantenido en los niveles de robustez de SIFT e incluso superiores cuando tenían lugar pequeños fallos debidos a los expuestos defectos de aplicación.
- Por otra parte, el método ha demostrado ser sensible al comportamiento del segmentador, especialmente en los bordes, por lo que la eficacia del método se verá limitada por éste factor.

En vista de lo expuesto en estas líneas, y a modo de conclusión global del proyecto:

Tras las pruebas y aplicaciones desarrolladas para testar el algoritmo implementado, se puede concluir que, a la luz de los resultados obtenidos en las mismas se ha conseguido crear un algoritmo capaz de mantener las propiedades fundamentales de SIFT y de corregir aquellas que suponían un defecto en ciertas aplicaciones.

Por otra parte, el método resulta sensible al funcionamiento del segmentador, especialmente en los bordes, por lo que se verá afectado negativamente por las condiciones de la imagen que hagan fallar al segmentador: ruido, bajas resoluciones o camuflajes.

8.2 Trabajo futuro

Habida cuenta del potencial presente en la idea implementada, resultan casi innumerables las posibilidades de trabajo futuro que se abren.

Algunas de las opciones de trabajo futuro planteadas, pero no desarrolladas por la duración finita de este proyecto se van a presentar categorizadas a continuación.

Las líneas de futuro principales en las que se considera que se podría trabajar a partir de esta idea son:

- Aplicaciones
- Método
- Eficiencia

Más detalladamente:

1) Aplicaciones:

Bien como descriptor principal o como complementario a otros ya implementados en ellas, se esperan obtener buenos resultados en las siguientes:

- *Object Cossegmentation*
- Reconocimiento de objetos
- Indexación de contenido
- Detección de duplicados

Por lo tanto, el uso de los descriptores *RegSIFT* o *groupedRegSIFT* podría suponer una mejora significativa para estas aplicaciones.

2) Método:

De cara a posibles mejoras sobre el propio algoritmo se pueden destacar:

- *Cambios del efecto introducido en los bordes de las regiones activas del descriptor*

Objetivo: La descripción llevada a cabo anulando la imagen en los bordes de la región activa da lugar a un método muy distintivo y característico. Sin embargo, esta descripción resulta absolutamente dependiente del buen comportamiento del segmentador. Es necesario reducir la sensibilidad a errores del segmentador.

Idea: Sería posible modificar el método desarrollado, pasando a anular el gradiente fuera de las regiones activas (en lugar de la imagen). Esta idea reduciría la dependencia del descriptor de los bordes,

manteniendo la finalidad de descripciones independientes del entorno pero dando lugar a *features* menos distintivas.

- *Regiones con variación de parámetros del segmentador*

Objetivo: Puesto que el modelo de diseño del método es idóneo salvo por la fuerte dependencia del segmentador, la otra posibilidad existente es la de mejorar su funcionamiento llevando a cabo algún tipo de ajuste paramétrico para optimizar sus resultados en cada caso.

Idea: Para conseguir un mejor ajuste del comportamiento del segmentador, cabría la posibilidad de implementar un método que generara “bolsas” de regiones, cada una de ellas obtenida en particiones resultantes del barrido de distintos parámetros de segmentación, de entre las que posteriormente pudiesen seleccionarse aquellas que mejor se ajustasen a los bordes buscados (por ejemplo buscando el *Máximo a posteriori* entre la evaluación de todas ellas).

3) Eficiencia:

Pese a presentar un tiempo de ejecución controlado para el elevado número de procesos que se le realizan para cada *feature*: $\sim 1 \text{ seg./frame}$, la eficiencia podría mejorarse mediante las siguientes soluciones:

- *Implementación óptima*

Si bien es cierto que se le ha dedicado tiempo a la implementación hasta lograr unos tiempos más que aceptables, siempre es posible mejorar aspectos técnicos del código.

- *Procesado en paralelo*

En la implementación final del algoritmo expuesto, la detección de puntos y de regiones se hace secuencialmente, pero dado que ambos procesos son independientes cabría la posibilidad de trabajarse con hilos, de forma que se pudieran hacer realmente ambos procesos en paralelo.

Referencias

- [1] LOWE, D.: "Object recognition from local scale-invariant features". In: *Computer Vision (ICCV) . The Proceedings of the Seventh IEEE International Conference on.*, 1999, S. 1150-1157 Vol. 2
- [2] BELONGIE, S.; MALIK, J. & PUZICHA, J.: "Shape matching and object recognition using shape contexts". In: *Pattern Analysis and Machine Intelligence (PAMI), IEEE Transactions on.*, 2002, S. 509—522 Nr 4
- [3] MIKOLAJCZYK, K.; LEIBE, B. & SCHIELE, B.: "Local features for object class recognition." In: *Computer Vision (ICCV). Tenth IEEE International Conference on.*, 2005, S. 1792-1799 Vol. 2
- [4] BROWN, M. & LOWE, D. G.: "Recognising panoramas." In: *Computer Vision (ICCV).*, 2003, S. 1218
- [5] SIVIC, J. & ZISSERMAN, A.: "Video Google: A text retrieval approach to object matching in videos." In: *Computer Vision (ICCV. Ninth IEEE International Conference on.*, 2003, S. 1470--1477
- [6] VAN DE SANDE, K. E.; GEVERS, T. & SNOEK, C. G.: "Evaluating color descriptors for object and scene recognition." In: *Pattern Analysis and Machine Intelligence (PAMI), IEEE Transactions on.*, 2010, S. 1582—1596 Nr. 9
- [7] SAHILLIOGLU, Y. & YEMEZ, Y.: "3d shape correspondence by isometry-driven greedy optimization." In: *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on.*, 2010, S. 453—458
- [8] HESS, R. & FERN, A.: "Discriminatively trained particle filters for complex multi-object tracking." In: *Computer Vision and Pattern Recognition (CVPR). IEEE Conference on.*, 2009, S. 240—247
- [9] TANG, C.-Y.; WU, Y.-L.; HOR, M.-K. & WANG, W.-H.: "Modified sift descriptor for image matching under interference." In: *Machine Learning and Cybernetics, International Conference on.*, 2008, S. 3294-3300
- [10] CUCCHIARA, R.; GRANA, C.; PICCARDI, M. & PRATI, A.: "Detecting moving objects, ghosts, and shadows in video streams." In: *Pattern Analysis and Machine Intelligence (PAMI), IEEE Transactions on .*, 2003, S. 1337—1342 Nr. 10
- [11] DARRELL, T.; MOGHADDAM, B. & PENTLAND, A. P.: "Active face tracking and pose estimation in an interactive room." In: *Computer Vision and Pattern Recognition (CVPR), Proceedings IEEE Computer Society Conference on.*, 1996, S. 67—72
- [12] FERGUS, R.; PERONA, P. & ZISSERMAN, A.: "Object class recognition by unsupervised scale-invariant learning." In: *Computer Vision and Pattern Recognition (CVPR), Proceedings. IEEE Computer Society Conference on.*, 2003, S. II--264
- [13] HARRIS, C. & STEPHENS, M.: "A combined corner and edge detector." In: *Alvey vision conference on.*, 1988, S. 50
- [14] MIKOLAJCZYK, K. & SCHMID, C.: "A performance evaluation of local descriptors." In: *Pattern Analysis and Machine Intelligence (PAMI), IEEE Transactions on.*, 2005, S. 1615—1630 Nr. 10
- [15] MIKOLAJCZYK, K. & SCHMID, C.: "Scale & affine invariant interest point detectors." In: *International journal of computer vision (ICCV), on.*, 2004, S. 63—86 Nr. 1
- [16] TUYTELAARS, T. & MIKOLAJCZYK, K.: "Local invariant feature detectors: a survey." In: *Foundations and Trends in Computer Graphics and Vision*, 2008, S. 177—280 Nr.3

- [17] MATAS, J.; CHUM, O.; URBAN, M. & PAJDLA, T.: "Robust wide-baseline stereo from maximally stable extremal regions." In: *Image and vision computing*, on., 2004, S. 761—767 Nr. 10
- [18] CANNY, J.: "A computational approach to edge detection." In: *Pattern Analysis and Machine Intelligence (PAMI)*, IEEE Transactions, on., 1986, S. 679—698 Nr. 6
- [19] BAY, H.; TUYTELAARS, T. & VAN GOOL, L.: "Surf: Speeded up robust features: Springer." *Computer Vision (ICCV)* on., 2006, S. 404--417
- [20] DORKÓ, G. & SCHMID, C.: "Selection of scale-invariant parts for object class recognition." In: *Computer Vision (ICCV), Proceedings. Ninth IEEE International Conference* on., 2003, S. 634—639
- [21] LEIBE, B. & SCHIELE, B.: "Scale-invariant object categorization using a scale-adaptive mean-shift search: Springer." *Pattern Recognition (CVPR)*, on., 2004, S. 145—153
- [22] LIU, J.; LUO, J. & SHAH, M.: "Recognizing realistic actions from videos "in the wild". In: *Computer Vision and Pattern Recognition (CVPR), IEEE Conference* on., 2009, S. 1996—2003
- [23] NOWAK, E.; JURIE, F. & TRIGGS, B.: "Sampling strategies for bag-of-features image classification: Springer." : *Computer Vision (ICCV)* on., 2006, S. 490—503
- [24] LOWE, D. G.: "Distinctive image features from scale-invariant keypoints." In: *International journal of computer vision (ICCV)* 2004, S. 91—110 Nr. 2
- [25] TA, D.-N.; CHEN, W.-C.; GELFAND, N. & PULLI, K.: SURFTrac: "Efficient tracking and continuous object recognition using local feature descriptors." In: *Computer Vision and Pattern Recognition,(CVPR) IEEE Conference* on., 2009, S. 2937—2944
- [26] HE, W.; YAMASHITA, T.; LU, H. & LAO, S.: "Surf tracking." In: *Computer Vision,(ICCV) IEEE 12th International Conference* on., 2009, S. 1586—1592
- [27] SERBY, D.; MEIER, E. & VAN GOOL, L.: "Probabilistic object tracking using multiple features." In: *Pattern Recognition(CVPR) Proceedings of the 17th International Conference* on., 2004, S. 184—187
- [28] DONOSER, M. & BISCHOF, H.: "Efficient Maximally Stable Extremal Region (MSER) Tracking." In: *Computer Vision and Pattern Recognition(CVPR), IEEE Computer Society Conference* on., 2006, S. 553-560
- [29] AGARWAL, A. & TRIGGS, B.: "3D human pose from silhouettes by relevance vector regression." In: *Computer Vision and Pattern Recognition (CVPR), Proceedings of the IEEE Computer Society Conference* on., 2004, S. II—882
- [30] MORI, G. & MALIK, J.: "Recovering 3d human body configurations using shape contexts." In: *Pattern Analysis and Machine Intelligence (PAMI), IEEE Transactions* on 2006, S. 1052—1062 Nr. 7
- [31] ZHU, G. & DOERMANN, D.: "Logo matching for document image retrieval." In: . . : *Document Analysis and Recognition(ICDAR), 10th International Conference* on., 2009, S. 606—610
- [32] JOHNSON, A. E. & HEBERT, M.: "Using spin images for efficient object recognition in cluttered 3D scenes." In: *Pattern Analysis and Machine Intelligence (PAMI), IEEE Transactions* on., 1999, S. 433--449 Nr. 5
- [33] VAN GOOL, L.; MOONS, T. & UNGUREANU, D.: "Affine/photometric invariants for planar intensity patterns: Springer." *Computer Vision (ICCV)*, on., 1996, S. 642--651
- [34] FLUSSER, J. & SUK, T.: "Pattern recognition by affine moment invariants." In: *Pattern recognition (CVPR)* 1993, S. 167—174 Nr. 1
- [35] BROWN, M. & LOWE, D. G.: "Invariant Features from Interest Point Groups." In: *British Machine Vision Conference (BMVC)*, 2002

- [36] K. MIKOLAJCZYK, "Detection of local features invariant to affine transformations," *PhD Thesis, Institut National Polytechnique de Grenoble, France, 2002.*
- [37] WU, Z.; KE, Q.; ISARD, M. & SUN, J.: "Bundling features for large scale partial-duplicate web image search." In: . : *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on., 2009, S. 25—32*
- [38] SCHAFFALITZKY, F. & ZISSERMAN, A.: "Multi-view matching for unordered image sets,; Springer." In: *Computer Vision (ICCV) on., 2002, S. 414--431*
- [39] LAZEBNIK, S.; SCHMID, C. & PONCE, J.: "A sparse texture representation using local affine regions." In: *Pattern Analysis and Machine Intelligence (PAMI), IEEE Transactions on., 2005, S. 1265—1278 Nr. 8*
- [40] ZHOU, H.; YUAN, Y. & SHI, C.: "Object tracking using SIFT features and mean shift." In: *Computer Vision and Image Understanding (CVPR), 2009 on., S. 345—352 Nr. 3*
- [41] MORI, G.; REN, X.; EFROS, A. A. & MALIK, J.: "Recovering human body configurations: Combining segmentation and recognition." In: *Computer Vision and Pattern Recognition (CVPR), Proceedings of the IEEE Computer Society Conference on., 2004, S. II—326*
- [42] REN, X. & MALIK, J.: "Learning a classification model for segmentation." In: *Computer Vision (ICCV), Proceedings. Ninth IEEE International Conference on., 2003, S. 10—17*
- [43] VEDALDI, A. & SOATTO, S.: "Quick Shift and Kernel Methods for Mode Seeking." In: *Proceedings of the European Conference on Computer Vision (ICCV) on., 2008*
- [44] TIBURZI, F.; ESCUDERO, M.; BESCÓS, J. & MARTÍNEZ, J. M.: "A ground truth for motion-based video-object segmentation." In: *Image Processing (ICIP) 15th IEEE International Conference on., 2008, S. 17—20*
- [45] BOWMAN, A. W. & AZZALINI, A.: "Applied smoothing techniques for data analysis." In: *Technometrics 41.3 1997*
- [46] KOKKINOS, I. & YUILLE, A.: "Scale invariance without scale selection." In: *Computer Vision and Pattern Recognition, (CVPR) IEEE Conference on., 2008, S. 1-8*
- [47] SENIOR, A.; HAMPAPUR, A.; TIAN, Y.-L.; BROWN, L.; PANKANTI, S. & BOLLE, R.: "Appearance models for occlusion handling." In: *Image and Vision Computing 2006, S. 1233 – 1243 Nr. 11*

Glosario

PRoIID	Points and Regions of Interest in Image Description
SIFT	Scale Invariant <i>Feature</i> Transform
SP	Superpíxeles
MSER	Maximally Stable Extremal Regions

Anexos

A Dataset del toy-example

A.1. Fondos Poco texturados



Figura 0-1: Imagen toy-example – I01



Figura 0-3: Imagen toy-example – I03



Figura 0-2: Imagen toy-example – I02

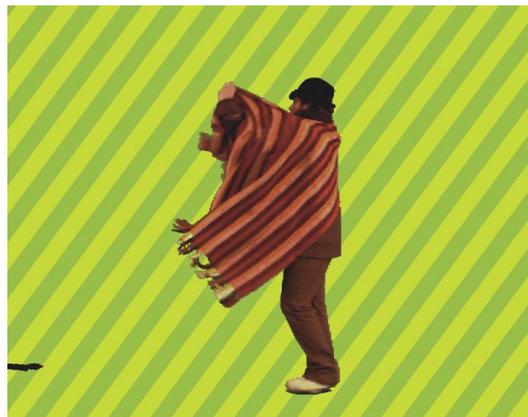


Figura 0-4: Imagen toy-example – I04

A.2. Fondos Altamente texturados

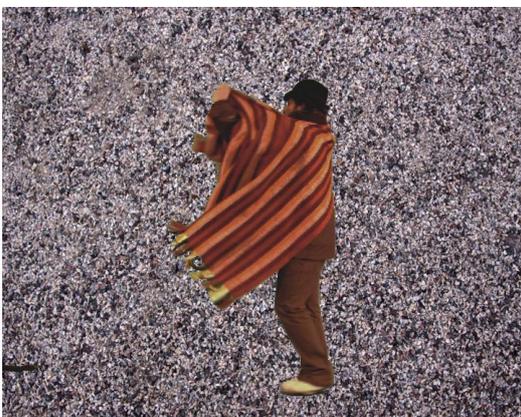


Figura 0-5 Imagen toy-example –IT01

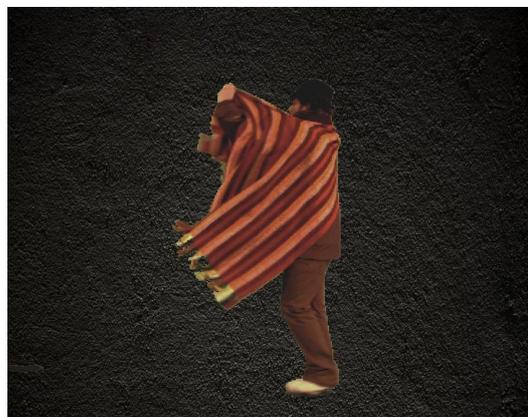


Figura 0-6: Imagen toy-example – IT02



Figura 0-7: Imagen toy-example – IT03

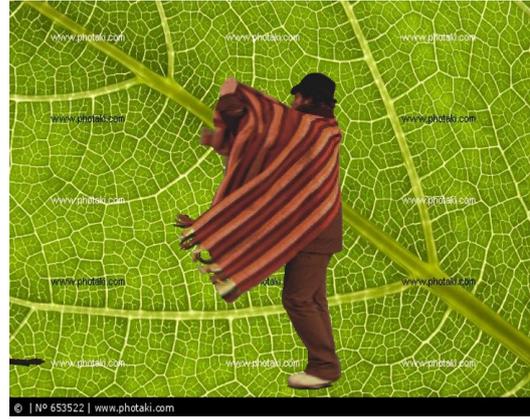
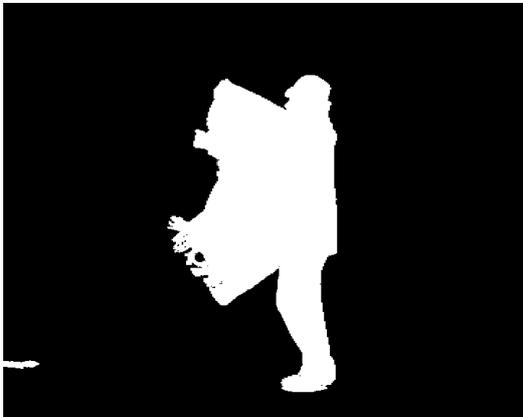


Figura 0-8: Imagen toy-example – IT04

A.3. Máscaras de ground-truth y cromas



0-9: Imagen toy-example – IT03



0-10: Imagen toy-example – IT04

B Dataset para análisis de propiedades

B.1. Imágenes de *Blur*



Figura 0-11: Imagen blur01



Figura 0-14: Imagen blur04



Figura 0-12: Imagen blur02



Figura 0-15: Imagen blur05



Figura 0-13: Imagen blur03



Figura 0-16: Imagen blur06

B.2. Imágenes de cambio de iluminación



Figura 0-17: Imagen light01



Figura 0-20: Imagen light04



Figura 0-18: Imagen light02



Figura 0-21: Imagen light05



Figura 0-19: Imagen light03



Figura 0-22: Imagen light06

B.3. Imágenes de cambio de punto de vista



Figura 0-23: Imagen viewpoint01



Figura 0-26: Imagen viewpoint 04



Figura 0-24: Imagen viewpoint 02



Figura 0-27: Imagen viewpoint 05



Figura 0-25: Imagen viewpoint 03



Figura 0-28: Imagen viewpoint 06

B.4. Imágenes cambio de escala + rotación



Figura 0-29: Imagen scl+rot01



Figura 0-32: Imagen scl+rot 04



Figura 0-30: Imagen scl+rot 02



Figura 0-33: Imagen scl+rot 05



Figura 0-31: Imagen scl+rot 03



Figura 0-34: Imagen scl+rot 06

C Dataset Logos Recognition



Figura 0-35: Imagen L1_1*



Figura 0-39: Imagen L2_2*



Figura 0-36: Imagen L1_2*



Figura 0-40: Imagen L2_3*



Figura 0-37: Imagen L1_3*



Figura 0-41: Imagen L3_1*



Figura 0-38: Imagen L2_1*



Figura 0-42: Imagen L3_2*



Figura 0-43: Imagen L3_3*



Figura 0-47: Imagen L5_1*



Figura 0-44: Imagen L4_1*



Figura 0-48: Imagen L5_2*



Figura 0-45: Imagen L4_2*



Figura 0-49: Imagen L5_3*



Figura 0-46: Imagen L4_3*

* Todas las imágenes marcadas han sido seleccionadas desde el buscador de imágenes de Google.

D Dataset Reconocimiento ante Oclusiones

D.1 Secuencia Objeto oclusivo Poco Texturado



Figura 0-50: Imagen OclusPT01



Figura 0-53: Imagen OclusPT04



Figura 0-51: Imagen OclusPT02



Figura 0-54: Imagen OclusPT05



Figura 0-52: Imagen OclusPT03



Figura 0-55: Imagen OclusPT06

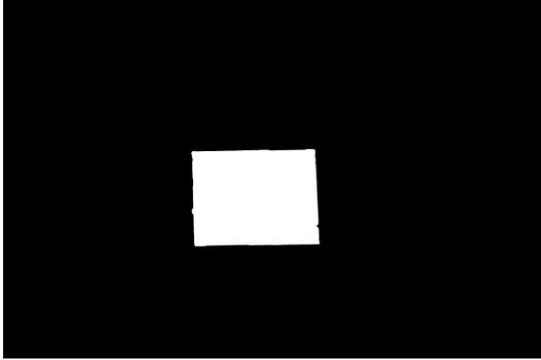


Figura 0-56: *ground-truth* OclusPT01

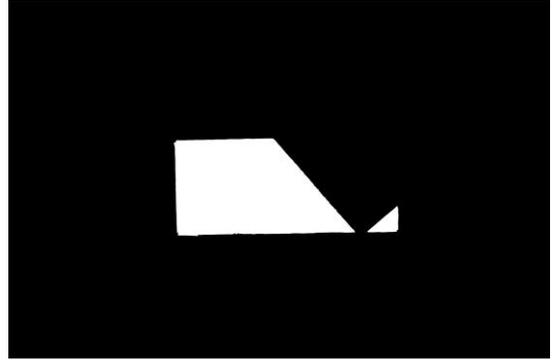


Figura 0-59: *ground-truth* OclusPT04

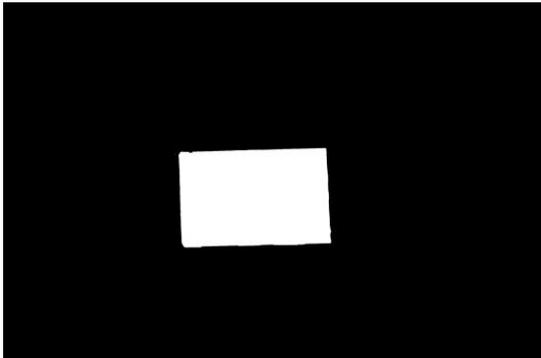


Figura 0-57: *ground-truth* OclusPT02

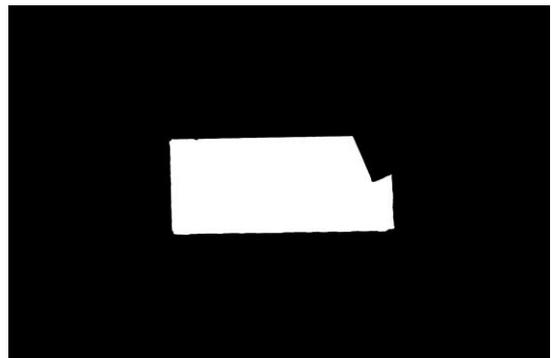


Figura 0-60: *ground-truth* OclusPT05

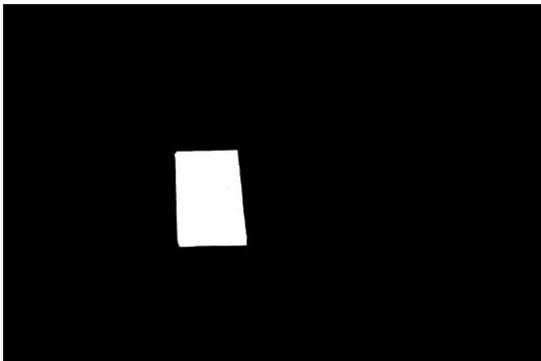


Figura 0-58: *ground-truth* OclusPT03



Figura 0-61: *ground-truth* OclusPT06

D.2 Secuencia Objeto oclusivo Altamente Texturado



Figura 0-62: Imagen OclusAT01



Figura 0-63: Imagen OclusAT02



Figura 0-64: Imagen OclusAT03



Figura 0-66: Imagen OclusAT05



Figura 0-65: Imagen OclusAT04



Figura 0-67: Imagen OclusAT06

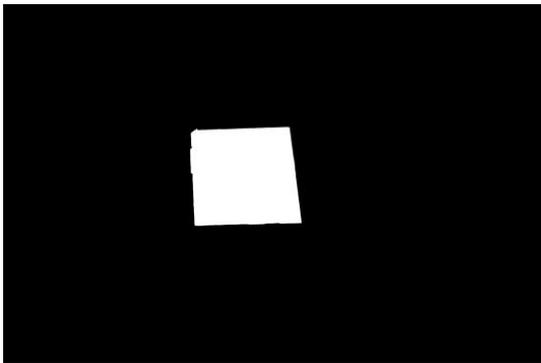


Figura 0-68: ground-truth OclusAT01

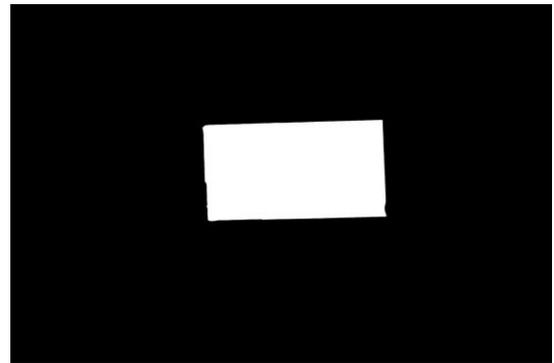


Figura 0-70: ground-truth OclusAT03

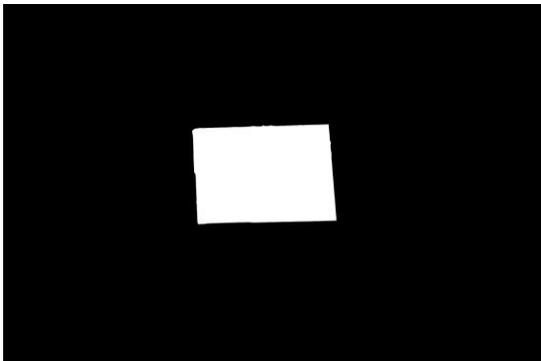


Figura 0-69: ground-truth OclusAT02

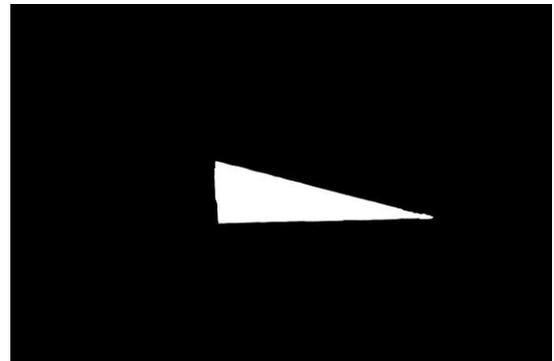


Figura 0-71: ground-truth OclusAT04

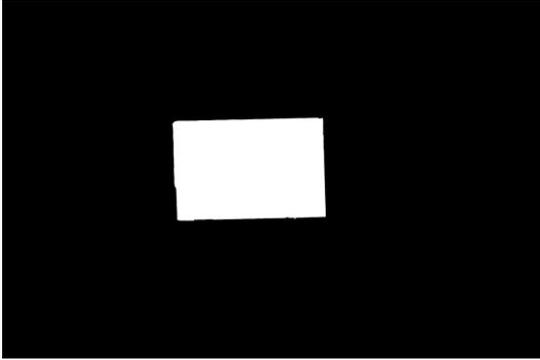


Figura 0-72: *ground-truth* OclusAT05

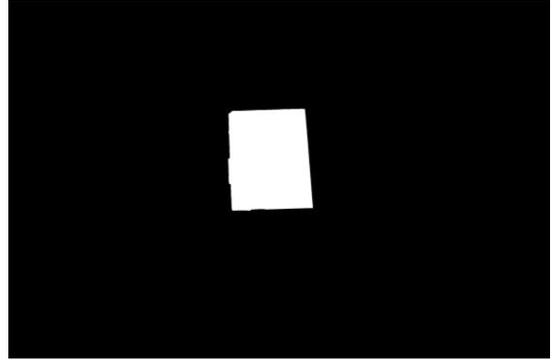


Figura 0-73: *ground-truth* OclusAT06

D.3 Secuencia Objeto doble oclusión y propiedades



Figura 0-74: Imagen OclusPDo01



Figura 0-77: Imagen OclusPDo04



Figura 0-75: Imagen OclusPDo02



Figura 0-78: Imagen OclusPDo05



Figura 0-76: Imagen OclusPDo03



Figura 0-79: Imagen OclusPDo06

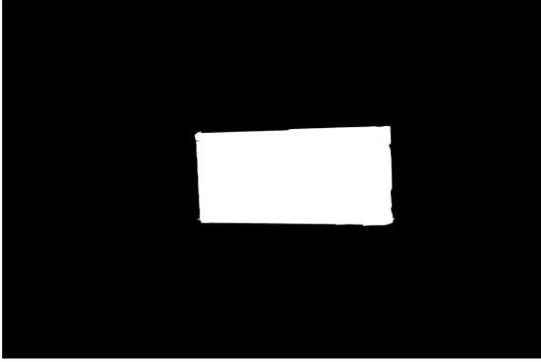


Figura 0-80: *ground-truth* OclusPD01

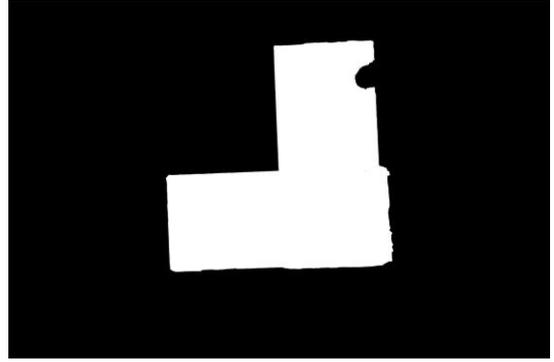


Figura 0-83: *ground-truth* OclusPD04

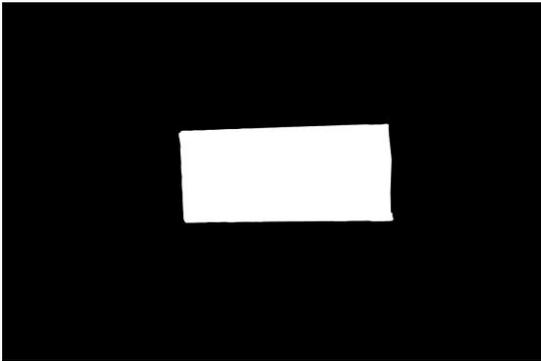


Figura 0-81: *ground-truth* OclusPD02

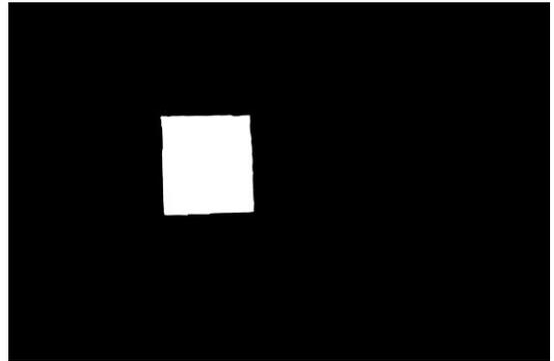


Figura 0-84: *ground-truth* OclusPD05

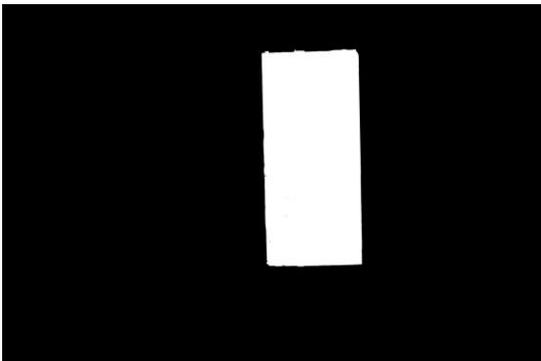


Figura 0-82: *ground-truth* OclusPD03

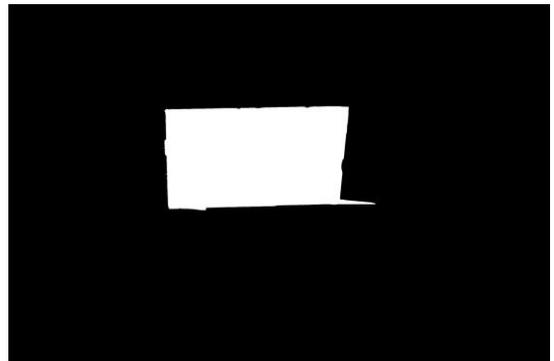


Figura 0-85: *ground-truth* OclusPD06

E Evolución del algoritmo

E.1 Introducción

El proceso del proyecto hasta la selección del método de SIFT como objetivo a mejorar, y la detección de sus defectos de aplicación siguió una evolución lineal. Sin embargo, el desarrollo del algoritmo hasta su configuración definitiva siguió una serie de estrategias de las cuales no todas llegaron a buen puerto.

Sus bases, objetivos, esquemas seguidos y motivos por los cuales fueron descartadas se detallarán en esta sección.

E.2 Línea de evolución

El proceso del desarrollo del algoritmo siguió cuatro etapas fundamentales, con una serie de pasos en ellas, como se observa en la Figura 0-86.

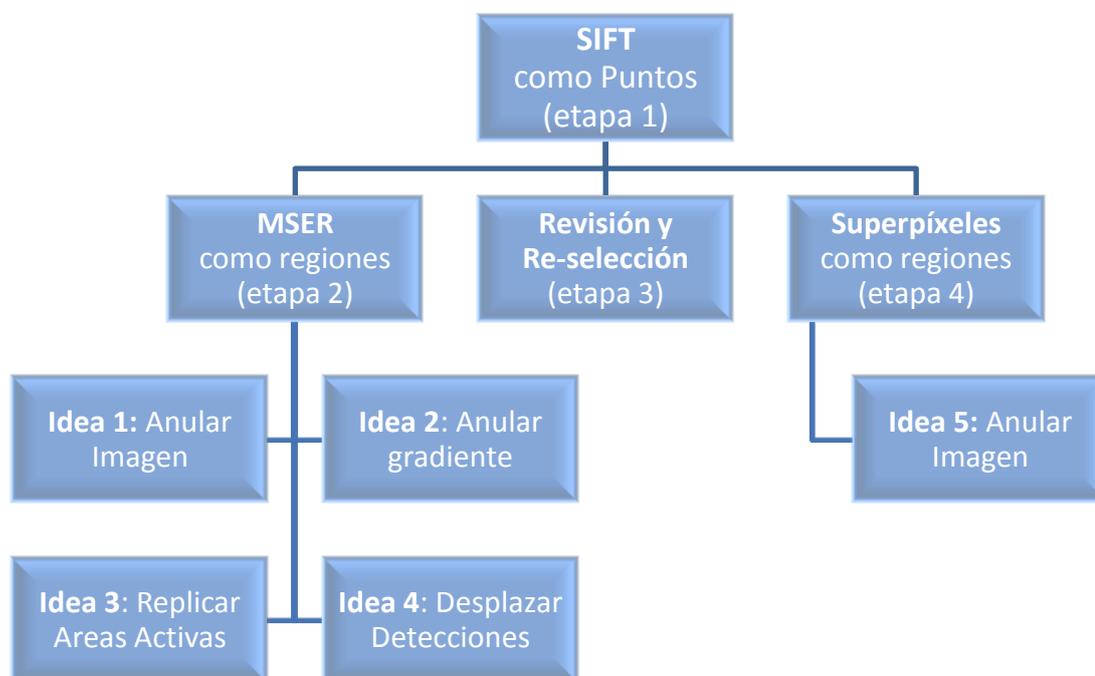


Figura 0-86: Esquema de implementaciones del algoritmo. Fuente: Propia

Las distintas ideas desarrolladas y su orden tienen una lógica que se expondrá en la sección E.3.

E.3 Desarrollo

E.3.1 Etapa 1: SIFT

Tras el estudio del estado del arte llevado a cabo en el Capítulo 2, la selección del algoritmo de SIFT se llevó a cabo según las conclusiones detalladas en la sección 2.4. Esta etapa no requiere pues comentario alguno sobre el proceso seguido más allá del especificado en dichos apartados.

E.3.2 Etapa 2: MSER

Tras el estudio llevado a cabo en el Capítulo 2 y la detección de los defectos asociados a aplicación en el Capítulo 4, se decidió desarrollar un algoritmo que introdujera en SIFT información de regiones.

Puesto que los mejores resultados a nivel de regiones se habían obtenido para las novedosas MSER [28], se decidió utilizar sus propiedades y funcionamiento para mejorar en los defectos de aplicación detectados para SIFT.

La estrategia de influencia en el método de SIFT fue la siguiente.

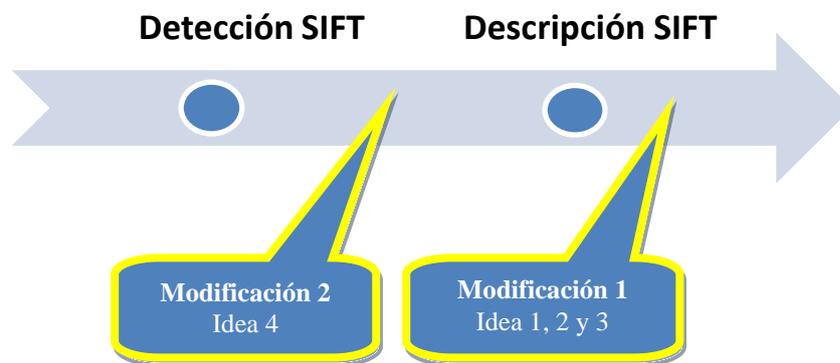


Figura 0-87: Línea de Modificaciones del Algoritmo. Fuente: Propia

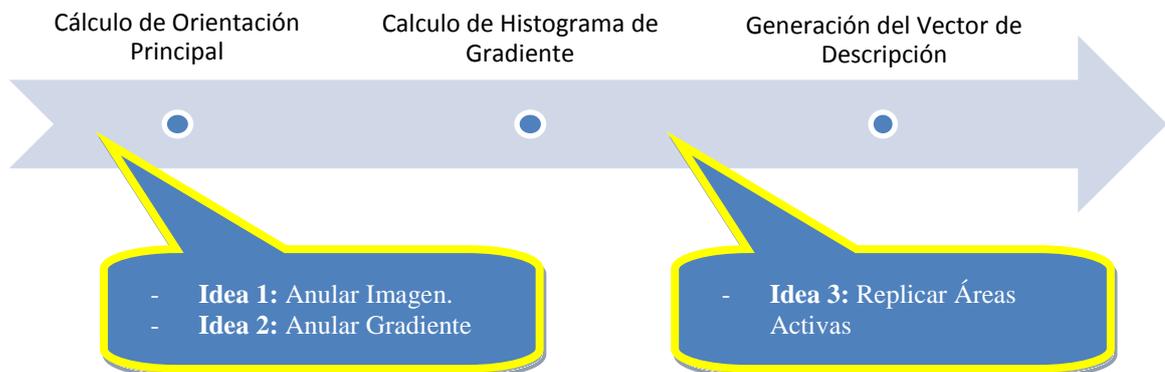


Figura 0-88: Detalle de Línea de Modificación 1. Fuente: Propia

1) Modificación 1

Como se ha podido observar en las anteriores Figura 0-87 y Figura 0-88, las Modificación 1 y sus distintas ideas, “etapas”, internas fueron llevadas a cabo únicamente sobre la etapa de descripción de SIFT. Cada una de ellas se presentará desglosada a continuación.

Idea 1: Anular Imagen

Síntesis

Dado que las MSER buscaban detectar regiones de Máxima Estabilidad, se consideró una buena idea anular la información que incorporase SIFT de fuera de las regiones MSER. Con esto, se lograría en principio que los puntos SIFT describiesen únicamente los objetos con las zonas más estables, evitando así incorporar información cambiante, que pudiese perjudicar a la descripción.

Resultado

Como resultado de la implementación de esta idea, se obtuvo un método que filtraba en exceso las detecciones de SIFT, reduciendo su número notablemente al igual que sus áreas de descripción, por lo que se obtenían pocos vectores de descripción y a su vez muy poco distintivos.

Descarte

A pesar de que las zonas MSER eran muy estables, no se lograba obtener vectores distintivos y suficientes para la descripción deseada, se decidió descartar este método.

Idea 2: Anular Gradiente

Síntesis

Puesto que el problema del punto anterior fue la excesiva reducción de la información que caracterizara los puntos, se decidió abordar la influencia desde otro punto de vista. En esta ocasión, se intervino en las zonas fuera de regiones MSER anulando el gradiente de la imagen, con la intención de que la información contenida en las regiones no quedará opacada por la simulación de un borde muy notable (resultado derivado de anular la imagen).

Resultado

Lejos de lograr que los vectores dejaran de ser poco distintivos, se observó que estos seguían siéndolo. Adicionalmente, seguían eliminándose excesivos puntos *post-detección* por localizarse completamente fuera de las MSER.

Descarte

Se concluyó que, los puntos no resultaban ser descritos poco distintivamente por la influencia introducida en los bordes de las regiones, sino porque las regiones MSER, por definición, incluían zonas muy estables, y por lo tanto de poca información respecto a su entorno.

Idea 3: Replicación de Regiones Activas

Síntesis

Vistos los malos resultados obtenidos de la intervención sobre la información de la imagen, se decidió intervenir una vez se habían realizado los cálculos de los histogramas de gradiente. Puesto que se seguía considerando positiva la idea de describir los elementos de la imagen únicamente mediante las zonas más estables, se decidió localizar los cuadrantes del área de descripción de cada punto SIFT que sí contenían MSER y replicar su información sobre los que no la contenían.

Resultado

Los vectores obtenidos tras este proceso resultaban ser altamente distintivos. Sin embargo, seguían localizándose numerosos puntos filtrados por no pertenecer a ninguna MSER. Por otra parte, pese a que la replicación se hacía en función de la orientación principal, se perdía en muchas ocasiones la invariancia a rotación por la inestabilidad de los bordes de las MSER.

Descarte

Puesto que la idea del proyecto no era únicamente lograr corregir los defectos de aplicación de SIFT, sino conseguirlo manteniendo las propiedades de SIFT lo más estables posibles, se decidió descartar esta implementación

2) *Modificación 2*

Una vez se había descartado la posibilidad de influir positivamente en la etapa de descripción de SIFT con las MSER, se pasó a contemplar la posibilidad de influir sobre el método en una etapa intermedia entre la detección de SIFT y su descripción. En este punto surgió la *Idea 4*.

Idea 4: Desplazar detecciones

Síntesis

Si bien es cierto que la detección de puntos SIFT era muy bien puntuada en los estudios comparativos de referencia [15], las regiones MSER también salían bien paradas en dichas comparativas. Puesto que la idea del algoritmo era conseguir modificar la descripción de SIFT para que esta se realizase únicamente con información estable y que no variará con el entorno cambiante se decidió lo siguiente.

Si las regiones MSER detectaban zonas de máxima estabilidad, desplazar las detecciones de SIFT hasta que partes de sus áreas de descripción contuvieran área MSER permitiría lograr el fin buscado. Una vez desplazadas se aplicarían la descripción de SIFT básica o alguna de las 3 ideas expuestas anteriormente.

Resultado

En esta ocasión, se conseguía evitar la desaparición excesiva de puntos. Sin embargo, pasando por alto las dificultades sobre la elección de hacia qué MSER desplazar cada punto y qué proporciones de área de MSER era suficiente conseguir, se observó que, el problema de descriptores poco distintivos o la pérdida de propiedades seguían presentes.

Descarte

El desplazamiento de puntos suponía la pérdida de propiedades y repetitibilidad logradas con los métodos de detección de SIFT, y puesto que éstas no eran compensadas con descripciones más robustas y distintivas se decidió descartar también esta idea.

E.3.3 Etapa 3: Revisión y Re-selección

Puesto que todos los resultados obtenidos para las pruebas realizadas sobre las distintas ideas expuestas en la sección E.3.2 fueron negativas, se decidió replantear la elección de las MSER como regiones que ayuden a la mejora del método de SIFT en sus defectos de aplicación.

Tras un estudio y análisis detallado de su funcionamiento en base a los resultados, y no a lo expuesto teóricamente por sus autores, se llegó a las siguientes conclusiones:

SIFT: Los puntos SIFT eran detectados en zonas inestables, de máxima variabilidad o singularidad. En su descripción incluían información del entorno, la cual, cuanto más elevada era, más distintiva y correctamente se extraía.

MSER: Detectaba zonas estables, sobre objetos o fondos. Sus detecciones, aunque repetitivas bajo una serie de transformaciones, resultaban ser extraordinariamente distintas en los bordes de la región para cada transformación. Adicionalmente, por su propia definición, evitaban ajustarse a los bordes si estos eran de alta intensidad.

Algoritmo Propio: Se busca que las detecciones sean igual que las que realiza SIFT, esto es, en puntos singulares de la imagen donde se puede describir mejor la misma. Por otra parte, se busca que la descripción que realiza sea también distintiva, y que incluya información suficiente aunque siendo capaz de diferenciar según sus regiones de origen.

En función de lo expuesto en las líneas anteriores, se decidió descartar el uso de las MSER por los siguientes motivos:

- La información que permitía describir resultaba poco distintiva y característica.
- Sus bordes eran inestables, por lo que no se podría realizar una descripción en función de su aspecto.
- Evitaba las zonas variables, donde se localizaban la mayoría de los puntos SIFT

Por lo tanto, aunque las MSER contenían, tal y como se deseaba, información estable de la imagen, se requería que estas se hubiesen ajustado a los bordes reales de la imagen de una manera estable e invariante, de tal manera que se pudieran lograr descripciones de puntos influenciados correctamente por estos contornos.

Descartados los métodos de detección de regiones tras los malos resultados obtenidos con las MSER, y con el nuevo requisito de regiones ajustadas a los bordes de las imágenes de manera estable, se llegó a la conclusión de que un segmentador y no un detector de regiones sería más útil para el propósito buscado. Así se llega a la etapa 4.

E.3.4 Etapa 4: Superpíxeles

Con las características buscadas de una técnica que permitiese la obtención de regiones, pero sin filtrar la información de bordes, se llegó a la conclusión de que se requería el uso de un segmentador de la máxima estabilidad posible.

Tal y como se expuso en la sección 5.2.2, los *Superpíxeles* resultaron ser los elegidos, y a la vista de los resultados obtenidos con ellos, la elección fue correcta.

PRESUPUESTO

- 1) **Ejecución Material**
 - Compra de ordenador personal (Software incluido)..... 2.000 €
 - Material de oficina 200 €
 - Total de ejecución material 2.200 €

- 2) **Gastos generales**
 - 16 % sobre Ejecución Material 352 €

- 3) **Beneficio Industrial**
 - 6 % sobre Ejecución Material 132 €

- 4) **Honorarios Proyecto**
 - 640 horas a 15 € / hora 9600 €

- 5) **Material fungible**
 - Gastos de impresión 60 €
 - Encuadernación 200 €

- 6) **Subtotal del presupuesto**
 - Subtotal Presupuesto 12060 €

- 7) **I.V.A. aplicable**
 - 21% Subtotal Presupuesto 2532,6 €

- 8) **Total presupuesto**
 - Total Presupuesto 14593,6 €

Madrid, Octubre de 2013

El Ingeniero Jefe de Proyecto

Fdo.: Fulgencio Navarro Fajardo
Ingeniero Superior de Telecomunicación

PLIEGO DE CONDICIONES

Este documento contiene las condiciones legales que guiarán la realización, en este proyecto, de una COMBINACIÓN DE EXTRACTORES Y DESCRIPTORES DE PUNTOS Y REGIONES DE INTERÉS. En lo que sigue, se supondrá que el proyecto ha sido encargado por una empresa cliente a una empresa consultora con la finalidad de realizar dicho sistema. Dicha empresa ha debido desarrollar una línea de investigación con objeto de elaborar el proyecto. Esta línea de investigación, junto con el posterior desarrollo de los programas está amparada por las condiciones particulares del siguiente pliego.

Supuesto que la utilización industrial de los métodos recogidos en el presente proyecto ha sido decidida por parte de la empresa cliente o de otras, la obra a realizar se regulará por las siguientes:

Condiciones generales

1. La modalidad de contratación será el concurso. La adjudicación se hará, por tanto, a la proposición más favorable sin atender exclusivamente al valor económico, dependiendo de las mayores garantías ofrecidas. La empresa que somete el proyecto a concurso se reserva el derecho a declararlo desierto.

2. El montaje y mecanización completa de los equipos que intervengan será realizado totalmente por la empresa licitadora.

3. En la oferta, se hará constar el precio total por el que se compromete a realizar la obra y el tanto por ciento de baja que supone este precio en relación con un importe límite si este se hubiera fijado.

4. La obra se realizará bajo la dirección técnica de un Ingeniero Superior de Telecomunicación, auxiliado por el número de Ingenieros Técnicos y Programadores que se estime preciso para el desarrollo de la misma.

5. Aparte del Ingeniero Director, el contratista tendrá derecho a contratar al resto del personal, pudiendo ceder esta prerrogativa a favor del Ingeniero Director, quien no estará obligado a aceptarla.

6. El contratista tiene derecho a sacar copias a su costa de los planos, pliego de condiciones y presupuestos. El Ingeniero autor del proyecto autorizará con su firma las copias solicitadas por el contratista después de confrontarlas.

7. Se abonará al contratista la obra que realmente ejecute con sujeción al proyecto que sirvió de base para la contratación, a las modificaciones autorizadas por la superioridad o a las órdenes que con arreglo a sus facultades le hayan comunicado por escrito al Ingeniero Director de obras siempre que dicha obra se haya ajustado a los preceptos de los pliegos de condiciones, con arreglo a los cuales, se harán las modificaciones y la valoración de las diversas unidades sin que el importe total pueda exceder de los presupuestos aprobados. Por consiguiente, el número de unidades que se consignan en el proyecto o en el presupuesto, no podrá servirle de fundamento para entablar reclamaciones de ninguna clase, salvo en los casos de rescisión.

8. Tanto en las certificaciones de obras como en la liquidación final, se abonarán los trabajos realizados por el contratista a los precios de ejecución material que figuran en el presupuesto para cada unidad de la obra.

9. Si excepcionalmente se hubiera ejecutado algún trabajo que no se ajustase a las condiciones de la contrata pero que sin embargo es admisible a juicio del Ingeniero Director de obras, se dará conocimiento a la Dirección, proponiendo a la vez la rebaja de precios que el Ingeniero estime justa y si la Dirección resolviera aceptar la obra, quedará el contratista obligado a conformarse con la rebaja acordada.

10. Cuando se juzgue necesario emplear materiales o ejecutar obras que no figuren en el presupuesto de la contrata, se evaluará su importe a los precios asignados a otras obras o materiales análogos si los hubiere y cuando no, se discutirán entre el Ingeniero Director y el contratista, sometiéndolos a la aprobación de la Dirección. Los nuevos precios convenidos por uno u otro procedimiento, se sujetarán siempre al establecido en el punto anterior.

11. Cuando el contratista, con autorización del Ingeniero Director de obras, emplee materiales de calidad más elevada o de mayores dimensiones de lo estipulado en el proyecto, o sustituya una clase de fabricación por otra que tenga asignado mayor precio o ejecute con mayores dimensiones cualquier otra parte de las obras, o en general, introduzca en ellas cualquier modificación que sea beneficiosa a juicio del Ingeniero Director de obras, no tendrá derecho sin embargo, sino a lo que le correspondería si hubiera realizado la obra con estricta sujeción a lo proyectado y contratado.

12. Las cantidades calculadas para obras accesorias, aunque figuren por partida alzada en el presupuesto final (general), no serán abonadas sino a los precios de la contrata, según las condiciones de la misma y los proyectos particulares que para ellas se formen, o en su defecto, por lo que resulte de su medición final.

13. El contratista queda obligado a abonar al Ingeniero autor del proyecto y director de obras así como a los Ingenieros Técnicos, el importe de sus respectivos honorarios facultativos por formación del proyecto, dirección técnica y administración en su caso, con arreglo a las tarifas y honorarios vigentes.

14. Concluida la ejecución de la obra, será reconocida por el Ingeniero Director que a tal efecto designe la empresa.

15. La garantía definitiva será del 4% del presupuesto y la provisional del 2%.

16. La forma de pago será por certificaciones mensuales de la obra ejecutada, de acuerdo con los precios del presupuesto, deducida la baja si la hubiera.

17. La fecha de comienzo de las obras será a partir de los 15 días naturales del replanteo oficial de las mismas y la definitiva, al año de haber ejecutado la provisional, procediéndose si no existe reclamación alguna, a la reclamación de la fianza.

18. Si el contratista al efectuar el replanteo, observase algún error en el proyecto, deberá comunicarlo en el plazo de quince días al Ingeniero Director de obras, pues transcurrido ese plazo será responsable de la exactitud del proyecto.

19. El contratista está obligado a designar una persona responsable que se entenderá con el Ingeniero Director de obras, o con el delegado que éste designe, para todo relacionado con ella. Al ser el Ingeniero Director de obras el que interpreta el proyecto, el contratista deberá consultarle cualquier duda que surja en su realización.

20. Durante la realización de la obra, se girarán visitas de inspección por personal facultativo de la empresa cliente, para hacer las comprobaciones que se crean oportunas. Es

obligación del contratista, la conservación de la obra ya ejecutada hasta la recepción de la misma, por lo que el deterioro parcial o total de ella, aunque sea por agentes atmosféricos u otras causas, deberá ser reparado o reconstruido por su cuenta.

21. El contratista, deberá realizar la obra en el plazo mencionado a partir de la fecha del contrato, incurriendo en multa, por retraso de la ejecución siempre que éste no sea debido a causas de fuerza mayor. A la terminación de la obra, se hará una recepción provisional previo reconocimiento y examen por la dirección técnica, el depositario de efectos, el interventor y el jefe de servicio o un representante, estampando su conformidad el contratista.

22. Hecha la recepción provisional, se certificará al contratista el resto de la obra, reservándose la administración el importe de los gastos de conservación de la misma hasta su recepción definitiva y la fianza durante el tiempo señalado como plazo de garantía. La recepción definitiva se hará en las mismas condiciones que la provisional, extendiéndose el acta correspondiente. El Director Técnico propondrá a la Junta Económica la devolución de la fianza al contratista de acuerdo con las condiciones económicas legales establecidas.

23. Las tarifas para la determinación de honorarios, reguladas por orden de la Presidencia del Gobierno el 19 de Octubre de 1961, se aplicarán sobre el denominado en la actualidad "Presupuesto de Ejecución de Contrata" y anteriormente llamado "Presupuesto de Ejecución Material" que hoy designa otro concepto.

Condiciones particulares

La empresa consultora, que ha desarrollado el presente proyecto, lo entregará a la empresa cliente bajo las condiciones generales ya formuladas, debiendo añadirse las siguientes condiciones particulares:

1. La propiedad intelectual de los procesos descritos y analizados en el presente trabajo, pertenece por entero a la empresa consultora representada por el Ingeniero Director del Proyecto.
2. La empresa consultora se reserva el derecho a la utilización total o parcial de los resultados de la investigación realizada para desarrollar el siguiente proyecto, bien para su publicación o bien para su uso en trabajos o proyectos posteriores, para la misma empresa cliente o para otra.
3. Cualquier tipo de reproducción aparte de las reseñadas en las condiciones generales, bien sea para uso particular de la empresa cliente, o para cualquier otra aplicación, contará con autorización expresa y por escrito del Ingeniero Director del Proyecto, que actuará en representación de la empresa consultora.
4. En la autorización se ha de hacer constar la aplicación a que se destinan sus reproducciones así como su cantidad.
5. En todas las reproducciones se indicará su procedencia, explicitando el nombre del proyecto, nombre del Ingeniero Director y de la empresa consultora.
6. Si el proyecto pasa la etapa de desarrollo, cualquier modificación que se realice sobre él, deberá ser notificada al Ingeniero Director del Proyecto y a criterio de éste, la empresa consultora decidirá aceptar o no la modificación propuesta.

7. Si la modificación se acepta, la empresa consultora se hará responsable al mismo nivel que el proyecto inicial del que resulta el añadirla.

8. Si la modificación no es aceptada, por el contrario, la empresa consultora declinará toda responsabilidad que se derive de la aplicación o influencia de la misma.

9. Si la empresa cliente decide desarrollar industrialmente uno o varios productos en los que resulte parcial o totalmente aplicable el estudio de este proyecto, deberá comunicarlo a la empresa consultora.

10. La empresa consultora no se responsabiliza de los efectos laterales que se puedan producir en el momento en que se utilice la herramienta objeto del presente proyecto para la realización de otras aplicaciones.

11. La empresa consultora tendrá prioridad respecto a otras en la elaboración de los proyectos auxiliares que fuese necesario desarrollar para dicha aplicación industrial, siempre que no haga explícita renuncia a este hecho. En este caso, deberá autorizar expresamente los proyectos presentados por otros.

12. El Ingeniero Director del presente proyecto, será el responsable de la dirección de la aplicación industrial siempre que la empresa consultora lo estime oportuno. En caso contrario, la persona designada deberá contar con la autorización del mismo, quien delegará en él las responsabilidades que ostente.