

Trabajo Fin de Grado

Grado de Ingeniería en Tecnologías Industriales

Comparación de puntos de interés en secuencias de imágenes visuales para localización y mapeado mediante robots móviles

Autora: Inés Herrero Fernández

Tutor: José Ramiro Martínez de Dios

Dpto. Ingeniería de Sistemas y Automática
Escuela Técnica Superior de Ingeniería
Universidad de Sevilla

Sevilla, 2020



Trabajo Fin de Grado
Grado en Ingeniería de Tecnologías Industriales

Comparación de puntos de interés en secuencias de imágenes visuales para localización y mapeado mediante robots móviles

Autora:

Inés Herrero Fernández

Tutor:

José Ramiro Martínez de Dios

Catedrático de la Universidad

Dpto. Ingeniería de Sistemas y Automática

Escuela Técnica Superior de Ingeniería

Universidad de Sevilla

Sevilla, 2020

Trabajo Fin de Grado: Comparación de puntos de interés en secuencias de imágenes visuales para localización y mapeado mediante robots móviles

Autora: Inés Herrero Fernández

Tutor: José Ramiro Martínez de Dios

El tribunal nombrado para juzgar el Proyecto arriba indicado, compuesto por los siguientes miembros:

Presidente:

Vocales:

Secretario:

Acuerdan otorgarle la calificación de:

Sevilla, 2020

El secretario del Tribunal

*A mi familia y quienes me apoyan.
A todos los que me cuidan más
allá de las estrellas.*

Agradecimientos

Me gustaría dedicar estas páginas a todas aquellas personas que han sido importantes en mi vida, no solo a mi familia y amigos, que han estado siempre para apoyarme, escucharme, aconsejarme y animarme a seguir adelante. Quiero agradecer también a todos aquellos, que sin saberlo, me han enseñado algo nuevo, y que con los años aprendo que sus palabras y consejos eran sabios. Además, también quiero agradecer a todos los obstáculos que se me han cruzado, porque sin ellos hoy no estaría donde estoy y no sabría apreciar todo lo bueno que tengo.

Inés Herrero Fernández

Sevilla, 2020

Resumen

En la actualidad, la robótica va creciendo día a día y extendiéndose cada vez más a diferentes campos. Cada vez es más fácil trabajar con ella, y por tanto su implementación en el uso cotidiano se puede ver más normal. Su uso puede ir desde la automoción, la telefonía, en la industria o incluso en los hogares. No solo es una tecnología que ayuda al ser humano, sino que incluso le facilita muchas labores.

En este trabajo se va a analizar la capacidad de otorgarle a los robots el sentido de la vista. Para ello, se van a analizar diferentes tecnologías que tratan de adaptar la visión humana en los robots. Se van a estudiar en diferentes vídeos, para obtener diferentes gráficas que se estudiarán. Se quiere observar la capacidad de análisis según las condiciones lumínicas del entorno, la capacidad de reconocer objetos y formas, el tiempo de respuesta, medir la eficiencia de cálculo ...

Este trabajo está estructurado en 5 capítulos, que se pasan a describir a continuación. En el primero de ellos, se hace una breve introducción al documento y se muestra la estructura. En el segundo, se muestra la documentación necesaria para poder entender y conocer las tecnologías que son objeto de estudio. En el capítulo tercero se introducen los diferentes análisis que se van a realizar y su utilidad, mostrándose los resultados de dichos estudios en el capítulo 4. Por último, se muestran las conclusiones y posibles aspectos de mejora en el capítulo final.

Índice

Agradecimientos	ix
Resumen	xi
Índice	xiii
Índice de Tablas	xiv
Índice de Figuras	xv
1 Introducción	1
1.1 <i>Introducción</i>	1
1.2 <i>Contexto</i>	1
1.3 <i>Estructura</i>	2
2 Estado del arte	3
2.1 <i>Introducción</i>	3
2.2 <i>SLAM visual</i>	3
2.3 <i>SIFT</i>	4
2.4 <i>SURF</i>	7
2.5 <i>ORB</i>	8
2.6 <i>RANSAC</i>	9
2.7 <i>Conclusiones</i>	10
3 Análisis realizados	11
3.1 <i>Introducción</i>	11
3.2 <i>Bags utilizados</i>	11
3.3 <i>Análisis del número de keypoints</i>	16
3.4 <i>Análisis de asociación de features</i>	17
3.5 <i>Análisis por condiciones de contraste</i>	17
3.6 <i>Análisis por costes computacionales</i>	18
3.7 <i>Conclusiones</i>	18
4 Discusión	19
4.1 <i>Introducción</i>	19
4.2 <i>Resultados del análisis del número de keypoints</i>	19
4.3 <i>Resultados del análisis de asociación de features</i>	42
4.4 <i>Resultados del análisis por condiciones de contraste</i>	49
4.5 <i>Resultados de costes computacionales</i>	53
4.6 <i>Comparación de extractores</i>	66
4.7 <i>Conclusiones</i>	67
5 Conclusiones y desarrollos futuros	69
5.1 <i>Conclusiones</i>	69
5.2 <i>Desarrollos futuros</i>	72
Referencias	73

ÍNDICE DE TABLAS

Tabla 4-1 : Porcentaje de repetición de keypoints	20
Tabla 4-2 : Porcentaje de falsos positivos en los puntos buenos	22
Tabla 4-3 : Porcentaje de keypoints buenos sin repetidos	23
Tabla 4-4 : Porcentaje de keypoints malos repetidos	25
Tabla 4-5 : Porcentaje de keypoints malos sin repetidos	25
Tabla 4-6 : Porcentaje de matches buenos	44
Tabla 4-7 : Porcentaje de matches malos	44
Tabla 4-8 : Porcentaje del número de fotos de cada vídeo respecto al total	54
Tabla 4-9 : Resumen de los resultados obtenidos por los extractores	66

ÍNDICE DE FIGURAS

Figura 2-1 : Definición de SLAM	3
Figura 2-2 : Esquema del algoritmo SLAM	4
Figura 2-3 : Distribución del histograma para la orientación en SIFT	6
Figura 2-4 : Método para calcular descriptor en SIFT	6
Figura 2-5 : Asignación de la orientación en SURF	8
Figura 2-6 : Ejemplo RANSAC	10
Figura 3-1 : Miniatura video puente	11
Figura 3-2 : Histograma vídeo puente	12
Figura 3-3 : Contraste vídeo puente	13
Figura 3-4 : Miniatura vídeo pista de baloncesto	13
Figura 3-5 : Histograma vídeo pista de baloncesto	14
Figura 3-6 : Contraste vídeo pista de baloncesto	14
Figura 3-7 : Miniatura vídeo pista de karts	15
Figura 3-8 : Histograma vídeo pista de karts	15
Figura 3-9 : Contraste vídeo pista de karts	16
Figura 3-10 : Ejemplo división imagen en cuadrículas	17
Figura 4-1 : Gráfica número de keypoints total	19
Figura 4-2 : Gráfica de keypoints repetidos	20
Figura 4-3 : Gráfica número total de keypoints, descartando repetidos	21
Figura 4-4 : Gráfica keypoints buenos	21
Figura 4-5 : Gráfica de keypoints buenos repetidos	22
Figura 4-6 : Gráfica de keypoints buenos sin repetidos	23
Figura 4-7 : Gráfica número de keypoints malos	24
Figura 4-8 : Gráfica keypoints malos repetidos	24
Figura 4-9 : Gráfica número de keypoints malos sin repetidos	25
Figura 4-10 : Gráfica número máximo de keypoints con respectivos buenos y malos	26
Figura 4-11 : Gráfica número máximo de keypoints buenos	27
Figura 4-12 : Gráfica número máximo de keypoints malos	27
Figura 4-13 : Gráfica número mínimo de keypoints y correspondiente buenos y malos (puente y karts)	28
Figura 4-14 : Gráfica número mínimo de keypoints y correspondiente buenos y malos (baloncesto)	28
Figura 4-15 : Gráfica número mínimo de keypoints malos	29

Figura 4-16 : Gráfica número mínimo de keypoints malos	29
Figura 4-17 : Gráfica distancia máxima entre keypoints	30
Figura 4-18 : Gráfica distancia mínima entre keypoints	31
Figura 4-19 : Gráfica distancia media máxima entre keypoints	31
Figura 4-20 : Gráfica distancia media mínima entre keypoints	32
Figura 4-21 : Gráfica distancia media SIFT vídeo puente	32
Figura 4-22 : Gráfica distancia media SIFT vídeo baloncesto	33
Figura 4-23 : Gráfica distancia media SIFT vídeo karts	33
Figura 4-24 : Gráfica distancia media SURF vídeo puente	34
Figura 4-25 : Gráfica distancia media SURF vídeo baloncesto	34
Figura 4-26 : Gráfica distancia media SURF vídeo karts	35
Figura 4-27 : Gráfica distancia media ORB vídeo puente	35
Figura 4-28 : Gráfica distancia media ORB vídeo baloncesto	36
Figura 4-29 : Gráfica distancia media ORB vídeo karts	36
Figura 4-30 : Número de keypoints total según cuadrante vídeo puente	37
Figura 4-31 : Número de keypoints total según cuadrante vídeo cancha baloncesto	38
Figura 4-32 : Número de keypoints total según cuadrante vídeo pista karts	38
Figura 4-33 : Número de keypoints buenos según cuadrante vídeo puente	39
Figura 4-34 : Número de keypoints buenos según cuadrante vídeo cancha baloncesto	39
Figura 4-35 : Número de keypoints buenos según cuadrante vídeo pista karts	40
Figura 4-36 : Número de keypoints malos según cuadrante vídeo puente	40
Figura 4-37 : Número de keypoints malos según cuadrante vídeo cancha baloncesto	41
Figura 4-38 : Número de keypoints malos según cuadrante vídeo pista karts	42
Figura 4-39 : Gráfica número de matches total	43
Figura 4-40 : Gráfica número de matches buenos	43
Figura 4-41 : Gráfica número de matches malos	44
Figura 4-42 : Número de matches total según cuadrante vídeo puente	45
Figura 4-43 : Número de matches total según cuadrante vídeo baloncesto	45
Figura 4-44 : Número de matches total según cuadrante vídeo karts	46
Figura 4-45 : Número de matches buenos según cuadrante vídeo puente	46
Figura 4-46 : Número de matches buenos según cuadrante vídeo cancha baloncesto	47
Figura 4-47 : Número de matches buenos según cuadrante vídeo pista karts	47
Figura 4-48 : Número de matches malos según cuadrante vídeo puente	48
Figura 4-49 : Número matches malos según cuadrante vídeo cancha baloncesto	48
Figura 4-50 : Número matches malos según cuadrante vídeo pista karts	49
Figura 4-51 : Contraste según cuadrantes vídeo puente	50
Figura 4-52 : Contraste según cuadrantes vídeo baloncesto	50
Figura 4-53 : Contraste según cuadrantes vídeo karts	51
Figura 4-54 : Gráfica número de keypoints según contraste vídeo puente	52

Figura 4-55 : Gráfica número de keypoints según contraste vídeo baloncesto	52
Figura 4-56 : Gráfica número de keypoints según contraste vídeo karts	53
Figura 4-57 : Gráfica número de fotos	54
Figura 4-58 : Gráfica tiempo calculo keypoints máximo (ms)	55
Figura 4-59 : Gráfica tiempo cálculo keypoints mínimo (ms)	55
Figura 4-60 : Gráfica tiempos cálculo keypoints SIFT video puente	56
Figura 4-61 : Gráfica tiempos cálculo keypoints SIFT video baloncesto	56
Figura 4-62 : Gráfica tiempos cálculo keypoints SIFT video karts	57
Figura 4-63 : Gráfica tiempos cálculo keypoints SURF video puente	57
Figura 4-64 : Gráfica tiempos cálculo keypoints SURF video baloncesto	58
Figura 4-65 : Gráfica tiempos cálculo keypoints SURF video karts	58
Figura 4-66 : Gráfica tiempos cálculo keypoints ORB video puente	59
Figura 4-67 : Gráfica tiempos cálculo keypoints ORB video baloncesto	59
Figura 4-68 : Gráfica tiempos cálculo keypoints ORB video karts	60
Figura 4-69 : Gráfica tiempo cálculo matches máximo (ms)	60
Figura 4-70 : Gráfica tiempo cálculo matches mínimo (ms)	61
Figura 4-71 : Gráfica tiempos cálculo matching SIFT video puente	61
Figura 4-72 : Gráfica tiempos cálculo matching SIFT video baloncesto	62
Figura 4-73 : Gráfica tiempos cálculo matching SIFT video karts	62
Figura 4-74 : Gráfica tiempos cálculo matching SURF video puente	63
Figura 4-75 : Gráfica tiempos cálculo matching SURF video baloncesto	63
Figura 4-76 : Gráfica tiempos cálculo matching SURF video karts	64
Figura 4-77 : Gráfica tiempos cálculo matching ORB video puente	64
Figura 4-78 : Gráfica tiempos cálculo matching ORB video baloncesto	65
Figura 4-79 : Gráfica tiempos cálculo matching ORB video karts	65

1 INTRODUCCIÓN

1.1 Introducción

La tecnología, al igual que la humanidad, avanza cada día a pasos agigantados buscando soluciones a problemas cotidianos o incluso llegando a lugares que antes no se habían pensado. Una de las herramientas más usadas es la robótica que cada día está más presente en la vida humana, desde construcción, investigación, incluso en los hogares con la aparición de robots de limpieza.

Si nos centramos en la robótica visual, es decir, aquella que se basa en la imagen percibida por cámaras y sensores, se nos abre un amplio mundo de posibilidades. A partir de la información obtenida por cámaras, se desarrollan diferentes técnicas para interpretar dichas imágenes. Gracias a esta interpretación, se pueden percibir objetos, definir el entorno captado por la cámara, detectar obstáculos etc.

1.2 Contexto

En la actualidad, una de las técnicas más empleadas en la robótica es SLAM (*Simultaneous Localization and Mapping*), que se explica con más profundidad en el apartado 2.2. En resumen, es una tecnología que apareció hace más de 30 años en la que se busca solución a problemas de mapeado y localización. Como se explica en su apartado, es una tecnología que se aplica en muchas áreas de conocimiento: salud, automovilismo, ciencia, audiovisual, robótica ... En concreto para este trabajo, se va a centrar en el SLAM visual para robots y la problemática de la localización.

La percepción de los robots es un ámbito que poco a poco se amplía, buscando siempre el objetivo de parecerse a la visión humana, incluso en ocasiones llegar a percibir datos que no son observables por el ser humano. Para el hombre es fácil reconocer formas: una esquina, un color, un cuadrado, un círculo... El trabajo está en otorgarles a los robots esta capacidad de visión, programándose para ello diferentes formas para dar solución a ello; estos métodos se les conoce como extractores visuales.

Una de las aplicaciones resultantes de la unión de la robótica y la visión es la robótica móvil, que se basa en el empleo de sensores y cámaras para decidir la acción del robot. Dependiendo de la finalidad y el campo de aplicación, se puede clasificar [1] la robótica móvil en:

- Robots industriales
- Seguridad
- Sanidad
- Investigación espacial
- Acción militar
- Educación
- Juguetes

En el ejemplo de un robot móvil implementado en la medicina, es fácil apreciar la importancia de la visión. En primer lugar, los datos obtenidos a través de la cámara deben ser claros, ya que en ocasiones se llega a sitios en los que el ser humano no es capaz de trabajar. En segundo, se ha de ser capaz de obtener estos datos de forma que la iluminación no sea un problema.

El objetivo de este trabajo es el estudio de varios extractores visuales (SIFT, SURF y ORB) para ver el funcionamiento de ellos en diversas situaciones de luminosidad, encontrar datos de manera rápida y eficaz,

perseguir la veracidad de datos, etc. Gracias a la ayuda de los extractores, se puede conseguir localizar características diferenciables de una imagen para después localizar objetos, enlazar imágenes y en un futuro, autolocalizar; uno de los objetivos que persigue SLAM.

1.3 Estructura

Este documento consta de 5 capítulos, cada uno de ellos con diferentes objetivos y propósitos. Se pasa a describir la organización de este trabajo:

1. **Introducción:**

Primer capítulo del trabajo, en el que se muestra a modo introductorio la estructura del trabajo y los objetivos de éste.

2. **Estado del arte:**

De manera introductoria, se comienza hablando de los métodos actuales de la robótica y la visión, introduciéndose el término de SLAM visual. Además, se presentan los diferentes extractores visuales que se han estudiado (SIFT, SURF y ORB) y sus principales características. Se incluye también un método estadístico que es clave en la verificación de los resultados obtenidos, llamado RANSAC.

3. **Análisis realizados:**

En este capítulo se introducen los diferentes videos usados para el estudio, resaltando las características de cada uno de ellos. Además, se muestran cuáles serán los estudios realizados y la utilidad de cada uno de ellos.

4. **Discusión:**

Capítulo en el que se muestran todas las gráficas y tablas realizadas en el estudio; en definitiva, los resultados obtenidos para los diferentes videos y para cada uno de los extractores.

5. **Conclusiones y desarrollos futuros:**

En el capítulo final del trabajo se resume todo lo acontecido en el capítulo anterior, tomándose conclusiones a raíz de los resultados obtenidos; mejores y peores extractores según el tiempo de computación, según el contraste de la imagen ... Además, se proponen diferentes soluciones a problemas obtenidos o incluso posibles estudios a realizar.

2 ESTADO DEL ARTE

2.1 Introducción

En el trabajo por modelar la visión humana en la robótica se han desarrollado diversos métodos y programas, conocidos como extractores visuales, que tienen como objetivo principal la búsqueda de características propias de una imagen como color, esquinas, curvaturas... que ayudarán a identificar objetos y formas. A estas características diferenciables se les llama keypoints o puntos clave.

Para entender un poco mejor los puntos clave se le invita a pensar en el típico juego de buscar diferencias entre dos imágenes o incluso buscar objetos en una escena. Si se jugara a buscar objetos, en primer lugar se identifican las características propias del objeto, tales como la curvatura, tamaño, color ... Una vez realizado, se pasa a buscar dichas características, o puntos clave, en la imagen original.

El fin de este trabajo es ver la capacidad de los diferentes detectores visuales en la búsqueda de puntos clave y enlace de imágenes, también conocido como matching. Por hacer una alusión al juego de buscar objetos, hacer matching sería el caso cuando se encuentra la característica que se estaba buscando.

A lo largo de este capítulo se muestra en primer lugar una tecnología muy conocida en la robótica llamada SLAM, su auge actual y la relación que tiene con este trabajo. A continuación, se desarrollan algunos detectores visuales y sus características principales que han sido estudiados para las diferentes pruebas: serán SURF, SIFT y ORB. Para finalizar, se muestra un método estadístico que se usa en muchas ocasiones en robótica, llamado RANSAC, muy útil para hacer predicciones y estimaciones sobre la autenticidad de datos, principalmente en el caso de matching.

2.2 SLAM visual

En el año 1986 [2] durante un congreso del IEEE en San Francisco, se plantea por primera vez un algoritmo para buscar soluciones en el posicionamiento y mapeado de escenarios; aparece el término SLAM (*Simultaneous Localization and Mapping*). Con esta tecnología lo que se busca es la creación de un mapa tridimensional que ayude a conocer el entorno y a partir de ahí localizar al dispositivo, idea que se ve reflejada en la figura 2-1 [3].



Figura 2-1 : Definición de SLAM

Actualmente, el uso de sensores y cámaras ayudan al primer paso de la localización; pero a pesar de ello queda lo más importante, que es el entendimiento del propio entorno o incluso la veracidad de los datos obtenidos por los sensores. Los errores producidos en las mediciones son no deseados, desde problemas climatológicos, acumulación de errores de medida entre otros. Debido a ello, se plantea la problemática de buscar una solución a ello, y por tanto crear un algoritmo que optimice los resultados y elimine aquello no deseado.

El funcionamiento de este algoritmo [4] comienza en primer lugar con la toma de datos a partir de sensores y cámaras, consiguiéndose puntos característicos del entorno. Con ellos, se comienza a estimar la posición usándose un EKF (*Extended Kalman Filter*) que se va actualizando con cada movimiento del robot.; dependiendo de si los puntos encontrados son nuevos o no, se vuelve a pasar de nuevo el filtro de Kalman. Se muestra a continuación en la figura 2-2 los pasos a seguir en el algoritmo:

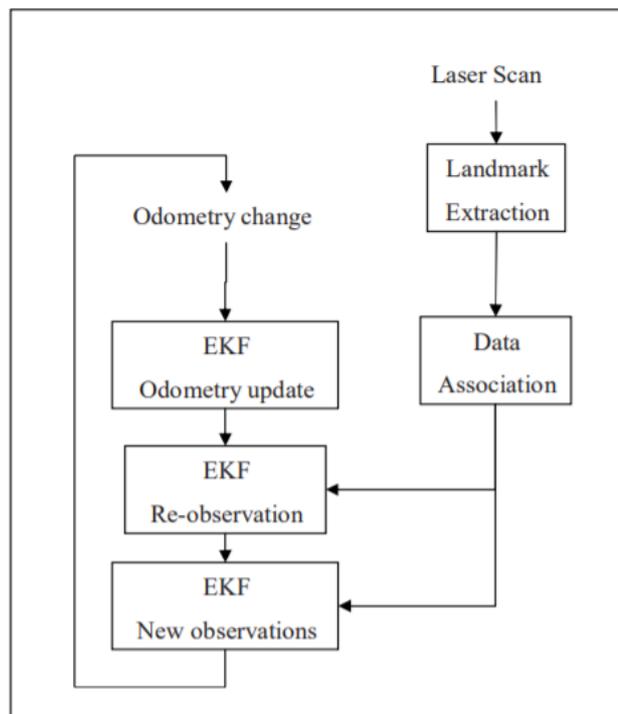


Figura 2-2 : Esquema del algoritmo SLAM

La implementación de esta tecnología se amplifica a varios campos, principalmente aquellos relacionados con la autonomía de robots y vehículos. Actualmente, el uso de cámaras y sensores implementados en la visión están a la orden del día, desde los sistemas de aparcamiento en los coches, pasando por los drones hasta incluso llegar a la visión propia de un robot que trabaja en el espacio.

Para poder aplicar SLAM, en primer lugar, hay que obtener puntos característicos y extraer de ellos información relevante que pueda resultar de interés: color, orientación, curvatura, tamaño etc. A estos puntos se les conocerá como keypoints y asociados a ellos tenemos sus características o descriptores. Una vez obtenidos dichos keypoints y sus descriptores se pretende enlazar con otra imagen para así reconocer objetos; el enlace de imágenes es lo que se conoce como matching. En los apartados siguientes se pasa a explicar diferentes extractores visuales cuyo objetivo es la búsqueda de puntos clave en imágenes y su posterior matching.

2.3 SIFT

En el año 2004, David G. Lowe publica un artículo llamado “*Distinctive Image Features from Scale-Invariant Keypoints*” [5] en el que presenta un nuevo algoritmo para extraer puntos clave de las imágenes y sacar sus descriptores; a este método se le conoce como SIFT que proviene del inglés *Scale Invariant Feature Transform*. Con este nuevo algoritmo se obtienen puntos clave y descriptores de manera que no sean función de la escala, iluminación ni rotación. Para llevar a cabo este método, se siguen una serie de pasos:

1. Escalado y espaciado:

El objetivo principal del escalado es la búsqueda de posibles keypoints, usándose para ello una estimación del laplaciano del gaussiano (*Laplacian of Gaussian*, LoG) para diferentes valores de σ o factor de escala; dicha aproximación será la diferencia del gaussiano (DoG).

Por definición, se define el laplaciano del gaussiano como:

$$LoG \triangleq \frac{x^2 + y^2 - 2\sigma^2}{\sigma^4} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \quad (2-1)$$

Al acercar una imagen de forma paulatina, se observa que las esquinas comienzan a perder curvatura y son más rectas; es por ello por lo que se usa LoG para diferentes valores de σ , considerándose dicho valor como un factor de escala. Debido al coste computacional del operador LoG, SIFT usa la aproximación de dicho método, conocido como DoG. Al tratarse de una diferencia, se calcula el gaussiano para diferentes σ para luego proceder a la resta.

Por definición, se define la diferencia de gaussiano como:

$$DoG \triangleq G\sigma_1 - G\sigma_2 = \frac{1}{\sqrt{2\pi}} \left(\frac{1}{\sigma_1} \exp\left(-\frac{x^2 + y^2}{2\sigma_1^2}\right) - \frac{1}{\sigma_2} \exp\left(-\frac{x^2 + y^2}{2\sigma_2^2}\right) \right) \quad (2-2)$$

Una vez obtenidas DoG, se pasa a localizar puntos extremos. Para ello, se compara con sus vecinos en la escala anterior y la siguiente. Gracias a la búsqueda de puntos extremos, se puede facilitar la posterior localización de keypoints.

2. Localización de keypoints:

Obtenidos los puntos extremos o posibles puntos clave, se debe definir y localizar exactamente su posición, usándose la serie de Taylor escala a escala. A medida que se avanza en el escalado se va calculando la intensidad del punto, descartándose aquellos con intensidad menos a 0.03; dicho factor puede ser modificado dentro de la librería y funciones ofrecidas por OpenCV [6].

Otro factor que calcular es el radio, necesario para la definición de esquinas. Para esta característica también se dispone de un umbral, con valor 10, que si no es superado se elimina dicho punto como posible candidato. Una vez realizado todo lo anterior, falta calcular la orientación del punto y unificar todas las características para obtener el descriptor del punto.

3. Asignación de orientación:

Gracias a este cálculo, se logran características invariantes a la rotación. Para ello, para cada punto L (x, y, σ) compuesto por sus coordenadas y su orientación se pasa a calcular la distancia entre píxeles.

Se define la ecuación de la distancia entre dos puntos como:

$$m(x, y) = \sqrt{[L(x + 1, y) - L(x - 1, y)]^2 + [L(x, y + 1) - L(x, y - 1)]^2} \quad (2-3)$$

Falta calcular el ángulo entre dos puntos como sigue:

$$\theta(x, y) = \tan^{-1} \left(\frac{L(x, y + 1) - L(x, y - 1)}{L(x + 1, y) - L(x - 1, y)} \right) \quad (2-4)$$

A partir de estas ecuaciones se calcula la distancia y la orientación con todos los píxeles que hay en

torno al keypoint, y se crea un histograma con dicha información. Este histograma está dividido en los 360 grados, que a su vez se divide en grupos de 10 grados; se obtiene por tanto 36 grupos. Para cada pareja de puntos se calcula la distancia, que será el valor del histograma para el valor de ángulo obtenido.

Una vez realizadas las operaciones con todos los puntos en torno al keypoint, se ha de observar en donde se produce picos de información entre el 80 y 100%; ese será el valor de la orientación [7]. En la figura 2-3 se muestra un ejemplo en el caso de obtener varios picos en el histograma.

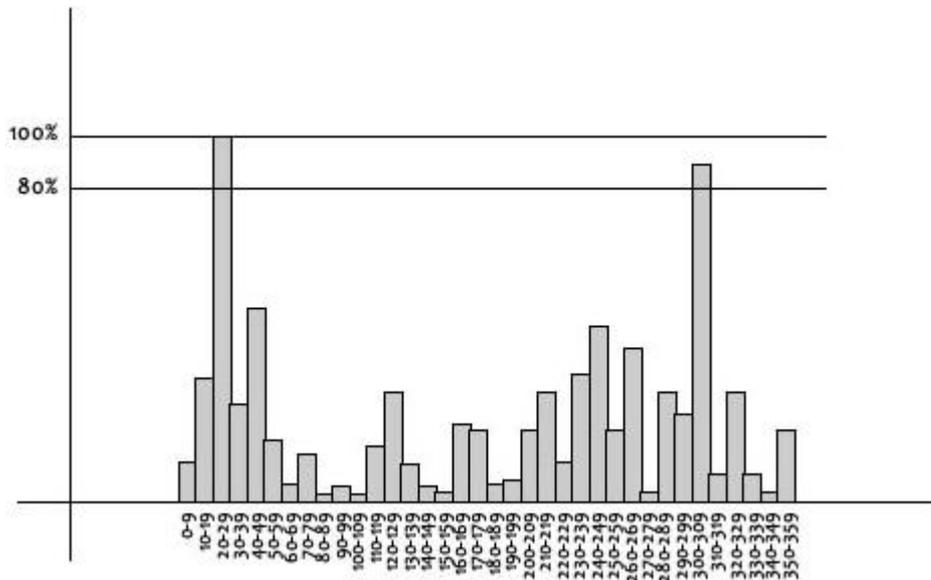


Figura 2-3 : Distribución del histograma para la orientación en SIFT

Si surgiera el caso de que hubiese más picos, se obtiene un nuevo keypoint con las mismas coordenadas y escala que el original, pero de orientación diferente; este aspecto se verá reflejado en los resultados que se podrán ver en el capítulo 4.

4. Descriptor del keypoint:

Para finalizar, se crea una identidad para keypoint; en definitiva, un vector de características que reúne todo lo necesario para poder identificarlo. Alrededor del punto se crea una cuadrícula de 16x16, dividiéndola a su vez en subbloques de 4x4. Para cada subbloque se vuelve a calcular y colocar en un histograma, dividido en 8 grupos, las distancias y la orientación. En total se obtienen 128 vectores que, una vez normalizados, serán las componentes del vector descriptor del keypoint; un ejemplo se muestra a continuación en la figura 2-4 [8], observándose como quedaría la cuadrícula en tono al punto clave.

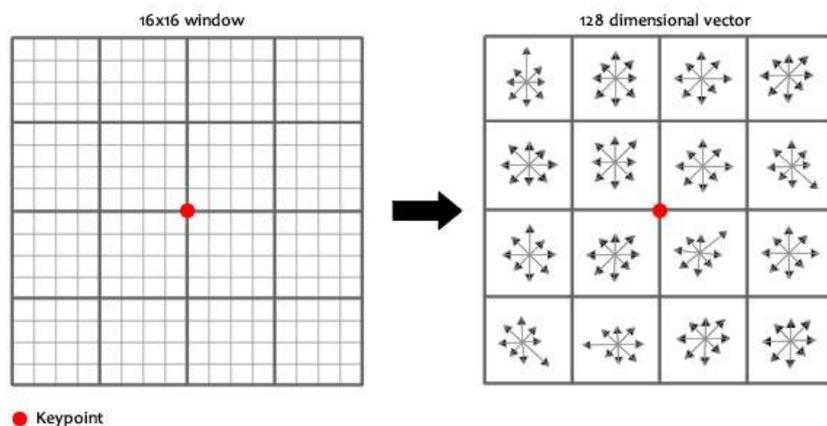


Figura 2-4 : Método para calcular descriptor en SIFT

2.4 SURF

En 2006, Herbert Bay, Tinne Tuytelaars y Luc Van Gool publican un documento, llamado “*SURF: Speeded Up Robust Features*” [9] en el que se presenta un nuevo detector de keypoints. Uno de los propósitos es conseguir un método de obtención de puntos más rápido que SIFT.

Como se comenta en el apartado anterior, el método SIFT usa una aproximación de LoG para cada valor de escala llamada DoG. Esta operación tiene un gran coste computacional, por lo que en el método SURF se pone remedio usando otra estimación de LoG llamada “*box filter*”. Se pasa a describir a continuación los pasos que se siguen en el método que pone título a este subapartado.

1. Detección de los puntos de interés:

En esta primera etapa el objetivo es buscar los posibles puntos clave, al igual que sucedía con SIFT. Ya que la gran diferencia de este extractor es la velocidad, en este caso se usa la matriz Hessiana y su determinante para los cálculos.

Sea un punto $p(x, y)$ dado por sus coordenadas, la matriz Hessiana $H(p, \sigma)$ según la escala se define como:

$$H(p, \sigma) = \begin{bmatrix} L_{xx}(p, \sigma) & L_{xy}(p, \sigma) \\ L_{xy}(p, \sigma) & L_{yy}(p, \sigma) \end{bmatrix} \quad (2-5)$$

Donde $L_{xx}(p, \sigma)$ es la convolución de segundo orden de la Gaussiana derivada $\frac{\delta^2}{\delta x^2} g(\sigma)$ de la imagen en el punto x ; de manera similar ocurre con el resto de los términos. A continuación, se aproximan las derivadas parciales, quedando el determinante de la matriz como:

$$\det(H_{approx}) = D_{xx}D_{yy} - (0.9 * D_{xy})^2 \quad (2-6)$$

El factor 0.9 es una aproximación que se encuentra explicada en el documento de SURF. Este determinante es interpolado en la escala y posición.

Con todo lo anterior se divide la imagen en diferentes escalas y octavas, eliminándose todos aquellos puntos que no cumplan la condición de máximo en un entorno de $3 \times 3 \times 3$.

2. Asignación de la orientación:

Para asignar la orientación, se debe calcular la respuesta de la onda de Haar (*Harr-wavelet*) para las direcciones x e y en torno a una región de tamaño $6s$, siendo s la escala a la que se encuentra el punto. El wavelet de Haar está descrito por:

$$\psi(t) = \begin{cases} 1; & 0 < t < \frac{1}{2} \\ -1; & \frac{1}{2} < t < 1 \\ 0; & \text{Otros casos} \end{cases} \quad (2-7)$$

Utilizándose imágenes integrales, se les aplica la onda anterior para las diferentes escalas y direcciones. Las respuestas obtenidas son ponderadas con una gaussiana con $\sigma=2.5$, y se representan tal y como se muestra en la figura 2-5 [10]:

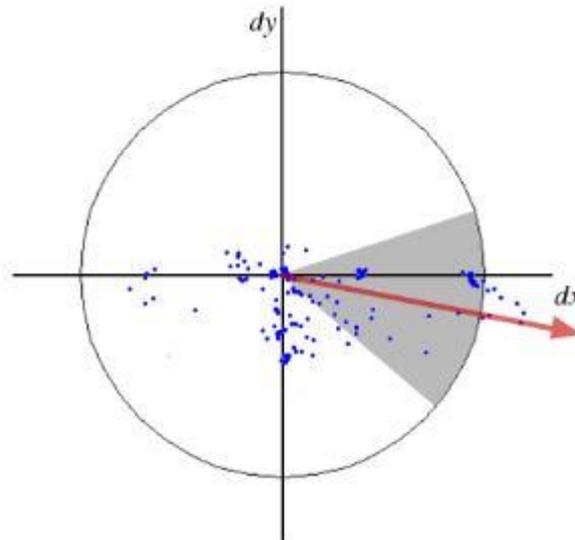


Figura 2-5 : Asignación de la orientación en SURF

Con dicha representación, se calcula en una sección móvil de tamaño $\pi/3$ la orientación dominante, que es resultado de la suma de los diferentes puntos representados.

3. Creación del descriptor SURF:

En el último paso, se crea una región cuadrada centrada en el keypoint de tamaño $20s$, siendo s la escala a la que se encuentra el punto. Esta región se divide en 4 cuadrados, que se dividen a su vez en 4, calculándose para cada uno de ellos la respuesta de la onda de Haar. Se crea un vector para cada subregión como sigue:

$$v = \left(\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y| \right) \quad (2-8)$$

Siendo dx y dy las respuestas de la onda de Haar para el eje correspondiente. Para cada punto se crea por tanto un descriptor con 64 componentes.

2.5 ORB

En 2011, se publica un nuevo artículo traído por Ethan Rublee, Vincent Rabaud, Kurt Konolige y Gary Brandiski, en el que se presenta una alternativa a los métodos SIFT y SURF. El nuevo extractor visual se presenta con el nombre de ORB que proviene de “*Oriented FAST and Rotated BRIEF*” [11], que es la unión de usar el detector de puntos FAST y el descriptor BRIEF. A diferencia de los métodos presentados anteriormente, este extractor es libre y se puede usar gratuitamente. A continuación, se pasa a explicar el funcionamiento del descriptor:

1. Detector de keypoints FAST:

Este tipo de detector busca en primer lugar los puntos FAST, calculándose para ello la intensidad umbral en torno a un punto central; en concreto se usa un radio de 9. A continuación, se usa Harris para así ordenar y descartar todos los candidatos que hayan salido en el paso anterior. Para la orientación, se usa de parámetro la intensidad medida en torno a un píxel, que se le pasa a llamar centroide. Esta medida, llamada intensidad del centroide, asume que las intensidades de las esquinas están compensadas con su centro, por lo que con ello se puede calcular un vector que vaya al centro, y por tanto la orientación del keypoint. Para ello, se calcula los momentos en x e y como sigue:

$$m_{pq} = \sum_{x,y} x^p y^q I(x,y) \quad (2-9)$$

Donde $I(x, y)$ es el valor en escala de grises, o lo que es lo mismo, la intensidad del punto

Con los momentos, se puede calcular el centroide como sigue:

$$C = \left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}} \right) \quad (2-10)$$

Con todo lo anterior se calcula la orientación del punto, ya que se conoce el centro y las esquinas, por tanto, sólo falta calcular la distancia del vector.

2. Creación del descriptor BRIEF:

El descriptor BRIEF (*Binary Robust Independent Elementary Features*) consiste en realizar una serie de test de intensidad en los pixeles ya encontrados anteriormente. Estos puntos están definidos por sus coordenadas (x, y) con las que se construye una matriz S de tamaño $2 \times n$, siendo n el número de pruebas realizadas. Debido a que el descriptor no funciona correctamente ante cambios en la rotación, se construye una nueva matriz S_θ , resultado de multiplicar S y la matriz de rotación, siendo θ la orientación del keypoint. Una vez obtenida la matriz S_θ , el método ORB discretiza los valores de los ángulos en intervalos de $2\pi/30$, colocándose los resultados en una tabla propia del método BRIEF. Mirando los resultados en la tabla, se escogerán como keypoints aquellos que sean consistentes en θ , es decir, su orientación.

2.6 RANSAC

El algoritmo RANSAC, que proviene de “*RANdom SAMple Concensus*” [12], es introducido por Martin A. Fishler y Robert C. Bolles en 1981. Este algoritmo es un método iterativo que se utiliza para estimar los parámetros matemático que modelan un conjunto de datos, es decir, darle una forma matemática a una serie de datos. Debido a ello, se puede comprobar si un conjunto de datos pertenece o no a la forma estimada, quitando aquellos que no pertenezcan para así evitar resultados erróneos; es similar a cuando se comprueba si unas coordenadas pertenecen a una función.

Este método estadístico es muy usado en visión computacional, principalmente para comprobar la autenticidad de los datos obtenidos. En concreto, para este proyecto se usa para verificar los matching realizados, ya que en ocasiones se realizan matching que no son correctos. Esto es debido a que el funcionamiento de enlazar puntos tiene como objetivo unir los puntos de la imagen original con la posterior, sin dejar ningún keypoint sin enlazar.

Se describe a continuación el funcionamiento del algoritmo, mostrándose además un ejemplo en la figura 2-6 [13]:

1. Se crea un subconjunto aleatorio de los datos originales, llamado “*inliers* hipotéticos”, y se crea un modelo a partir de ellos.
2. Se comprueba con el resto de los datos si se ajustan al modelo que se acaba de diseñar.
3. Se repiten los pasos anteriores, creándose otros modelos. Aquel modelo que consiga ajustar el mayor número de puntos será el elegido.

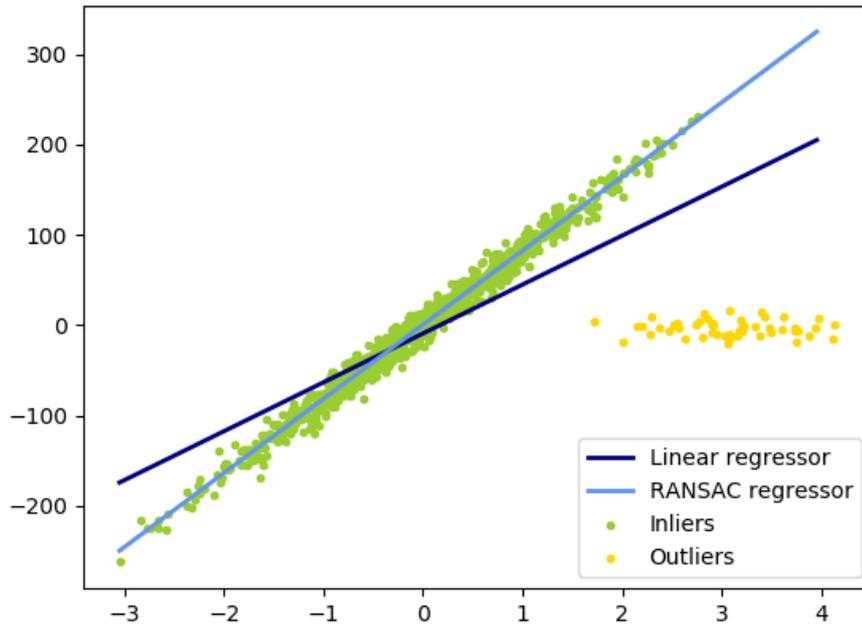


Figura 2-6 : Ejemplo RANSAC

Como se aprecia en la ilustración anterior, el objetivo es modelar la función que represente lo mejor posible los datos que se tienen; en este caso sería la línea celeste. Todos los puntos que se muestran serán los datos obtenidos en las diferentes pruebas, observándose algunos que pertenecen al modelo (*inliers*) y otros no (*outliers*). Lo que se pretende es quitar los *outliers* ya que, si se tuviesen en cuenta, los resultados estadísticos no se ajustarían correctamente a la realidad.

2.7 Conclusiones

La introducción a los diferentes extractores visuales presentados en el capítulo sirve para el conocimiento de su funcionamiento. El objetivo ahora es conocer la respuesta y eficiencia en diferentes situaciones; respuesta temporal, eficiencia en el matching, características extraídas ... Para ello, se cuenta con tres videos a estudiar, cada uno con características diferentes, que se analizarán para cada uno de los extractores. Además, para verificar la autenticidad de los keypoints encontrados y los matching realizados se usa el método estadístico de RANSAC, evitando así los datos que puedan influir negativamente en los resultados, es decir, quitando los outliers del estudio.

3 ANÁLISIS REALIZADOS

3.1 Introducción

En este capítulo se comienza a entrar en la materia de investigación de este trabajo. En primer lugar, se presentan los diferentes videos o bags que se han utilizado para el estudio. Para ello, se muestran diferentes características de ellos tales como su duración, histograma acumulado etc. Una vez presentados los bags utilizados, se pasa a explicar los diferentes análisis que se han llevado a cabo. Para cada uno de los análisis realizados, se muestra el interés por el cual se han llevado a cabo, así como cuáles han sido los estudios que se han hecho.

Para las diferentes pruebas, se ha trabajado con un código programado en C++ implementado en ROS (*Robot Operating System*) [14]. Este sistema operativo cuenta con diferentes herramientas y librerías que son útiles en el desarrollo de este proyecto; en concreto, se usa la librería OpenCV (*Open Computer Vision*). Con esta librería se abre un amplio mundo de posibilidades en la visión, desde trabajar con imágenes, videos en tiempo real o incluso extraer características que interesen; estos extractores visuales se presentaron en el capítulo 2.

En este proyecto se han estudiado los videos para cada uno de los extractores, obteniéndose para cada uno diferentes resultados. Los datos obtenidos tras los análisis se muestran en el capítulo siguiente, recogidos en tablas que se pasaran a analizar.

3.2 Bags utilizados

Para el estudio de los diferentes extractores (SIFT, SURF y ORB), se ha contado con tres videos. Como se ha trabajado en ROS, los videos utilizados serán un formato específico usado en este sistema que se conoce como bag. Las cámaras que se han usado para las grabaciones son estereoscópicas, es decir, se obtienen grabaciones tanto para la derecha e izquierda que después pueden formar una imagen 3D. Los videos han sido grabados en torno a las instalaciones de la escuela, pasándose a describir cada uno de ellos a continuación:

1. Puente:

Duración: 1'32''

Fecha grabación: 24 de agosto de 2017.

Miniatura del video: En la figura 3-1 se muestra una imagen tomada del video.



Figura 3-1 : Miniatura video puente

Tamaño de cada imagen: 672x376 píxeles.

Número de imágenes: 3049.

Histograma tonal:

Se muestra a continuación en la figura 3-2 el histograma calculado para todo el video tomado en escala de grises.



Figura 3-2 : Histograma vídeo puente

Si se divide el histograma en 3 zonas para diferenciar entre valores oscuros, medio y claros, cada sección estará formada por 85 valores. Como se puede apreciar, se distinguen dos picos: uno en la parte más oscura (entre 0 y 85) y otro en la parte central (entre 85 y 170). Para la parte de valores claros (entre 170 y 255) se obtienen valores constantes, es decir, no es apreciable el número de píxeles claros en comparación con los tonos medios y oscuros.

En conclusión, este video tiene el contraste bien proporcionado, no existe una zona dominante tonal en comparación con las otras.

Contraste:

Para cada una de las imágenes que componen el video se ha calculado el contraste, resultado la gráfica que se muestra en la figura 3-3. El cálculo se ha realizado obteniéndose la función de distribución del histograma, observándose la respuesta entre el 10 y 90%; la diferencia entre estos dos valores será el contraste.

Se define la función de distribución de una variable aleatoria discreta [15] como:

$$F(x) = \sum_{x_i \leq x} f(x_i) \quad (3-1)$$

El caso que se estudia es el número de píxeles que tienen cierto tono de gris. Es por tanto que la suma del histograma da como resultado el número total de píxeles presentes en la imagen. Por ello, cada punto de la gráfica de distribución será calculado como el valor del histograma dividido entre el número total de píxeles. Dada una función de distribución F se cumple que $0 \leq F(x) \leq 1$, por lo que cuando se desea calcular la respuesta entre el 10 y 90% se obtiene como la diferencia entre los puntos 0.9 y 0.1. En la figura que se presenta a continuación (3-3) se muestra el valor obtenido del contraste para el vídeo puente.



Figura 3-3 : Contraste vídeo puente

Se comprueba con la imagen anterior que los valores del contraste están repartidos en su rango, es decir, las imágenes están contrastadas. Como se observó en su histograma correspondiente (3-2), la característica resaltante del video es que predominan las zonas oscuras y con colores medios. Se aprecia que existen zonas claras o con mucho contraste, aunque no son tan llamativas como ocurre en el caso de valores sombríos, donde se observa un pico en torno a 45, dándose en más de 120 fotos.

2. Pista de baloncesto:

Duración: 2'52''

Fecha grabación: 7 de diciembre del 2017

Miniatura del video: En la figura 3-4 se muestra una imagen representativa del video.



Figura 3-4 : Miniatura vídeo pista de baloncesto

Tamaño de cada imagen: 672x376 píxeles

Número de imágenes: 5984

Histograma tonal:

Se muestra a continuación en la figura 3-5 el histograma de las imágenes en escala de grises que componen el video.

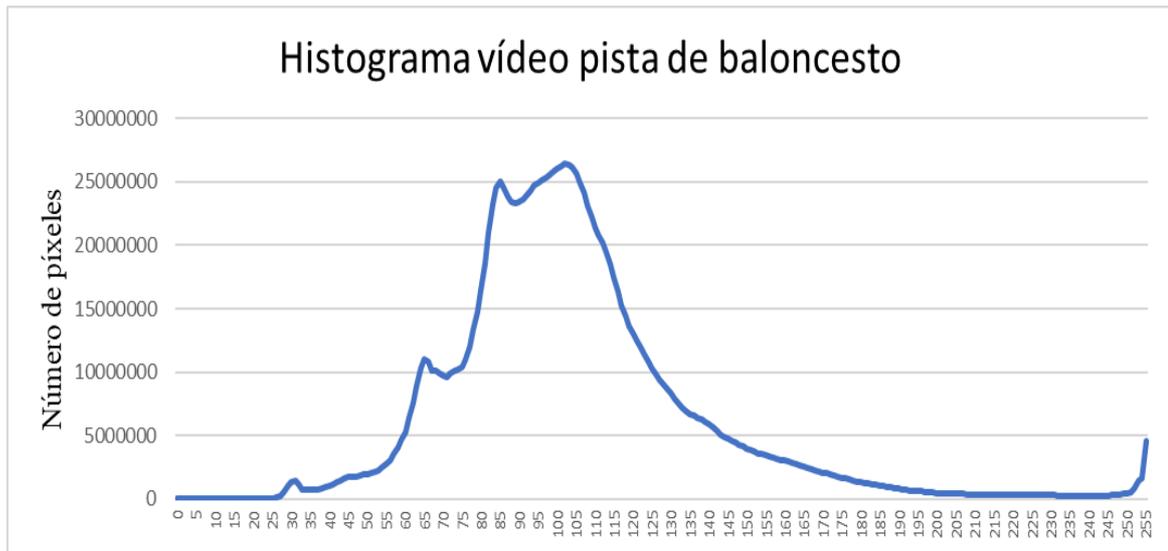


Figura 3-5 : Histograma vídeo pista de baloncesto

A diferencia del vídeo anterior, en esta ocasión se aprecia que la mayoría de los píxeles se concentran en la zona central de la gráfica, obteniéndose valores constantes en las zonas extremas. Esto quiere decir que el video que se está analizando está poco contrastado, ya que sólo se obtienen valores dominantes en una de las zonas.

Contraste:

En la figura 3-6 se muestra el contraste que se obtiene tras analizar cada foto que compone el vídeo grabado en la pista de baloncesto.



Figura 3-6 : Contraste vídeo pista de baloncesto

Una de las características más resaltables de la imagen anterior es que se aprecia un claro allanamiento de la curva en la zona de píxeles blancos, por lo que se deduce que las imágenes están subexpuestas. El conjunto de datos se centra en el sector oscuro y de tonos medios, observándose un pico en el valor 95, con cerca de 260 fotos.

3. Pista de karts: Duración: 2'22''

Fecha grabación: 26 de julio de 2018

Miniatura del video: A continuación, en la figura 3-7, una muestra del video que se está describiendo.

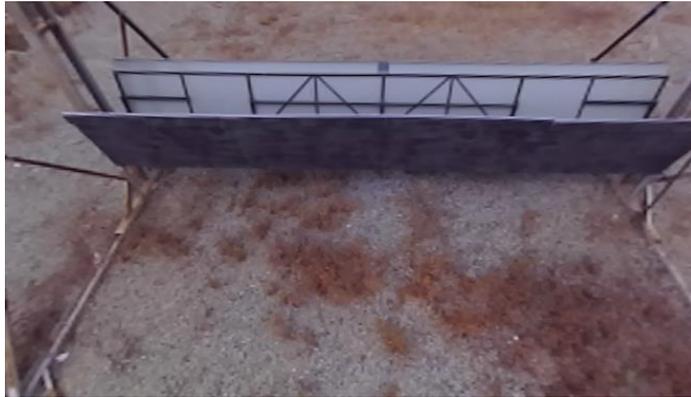


Figura 3-7 : Miniatura vídeo pista de karts

Tamaño de cada imagen: 672x376 píxeles

Número de imágenes: 334

Histograma tonal:

En la imagen 3-8 se muestra una representación gráfica del histograma del vídeo.



Figura 3-8 : Histograma vídeo pista de karts

En este histograma se aprecia que las imágenes no están sobreexpuestas, es decir, no se ven picos en la zona clara de la gráfica (valores entre 170 y 255). El pico se ve claramente en la zona central, aunque existe un acumulamiento de datos en la zona comprendida entre 0 y 85, principalmente en el término de la zona. Debido a que no se aprecian más picos notorios, se llega a la conclusión de que se está ante el caso de un video poco contrastado.

Contraste:

Se presenta en la figura 3-9 el resultado obtenido tras calcular el contraste del video titulado “pista de karts”



Figura 3-9 : Contraste vídeo pista de karts

Como ya se vaticinó en el histograma, el video se encuentra subexpuesto, ya que los valores del contraste en la zona de píxeles claros son prácticamente nulos. Se produce un cúmulo de información en la zona comprendida entre 25 y 135 aproximadamente, que se asocia con los tonos oscuros y medios. Se observan varios picos, el más notorio en la zona oscura con un valor de 40 en casi 20 fotos. Otro pico remarcable está en el valor 103, que pertenece a los valores medios de color, aunque a pesar de ello en esta zona hay varios picos y no destacan de manera significativa en la gráfica global.

3.3 Análisis del número de keypoints

En este apartado se muestran las diferentes mediciones que se han realizado con los keypoints obtenidos, es decir, los puntos claves de la imagen para cada vídeo y cada extractor. Estas pruebas son:

- Número total de keypoints obtenidos.
- Número total de keypoints, descartándose los repetidos; ocurre con mayor probabilidad en SIFT y en ocasiones con ORB. La explicación de este fenómeno se encuentra en el apartado 2.3, cuando se habla de la orientación del punto.
- Número total de keypoints repetidos, en el caso de los extractores SIFT y ORB.
- Diferenciación de keypoints buenos y malos, es decir, aquellos con los que se obtiene un match correcto o incorrecto. Con estos resultados se obtiene además el porcentaje, tanto de buenos como de malos.
- Número de keypoints máximos que se obtienen para una foto. Además, con esta característica se añade también el cálculo de puntos máximos obtenidos correctos e incorrectos (que realizan match bueno o malo).
- Número de keypoints mínimos que se obtienen para una foto. Al igual que el anterior, se obtienen los correctos y los incorrectos.
- Distancia máxima y mínima medida entre los keypoints. Con esta característica se busca conocer si los puntos están muy espaciados dentro de la imagen o si se concentran en una zona.
- Distancia media entre los keypoints; se obtiene la máxima, la mínima y la media. Debido a la variedad en la distribución de los puntos, se ha optado por medir las distancias medias.

Otro estudio realizado se ha basado en dividir la imagen en 12 cuadrantes, tal y como se muestra en la figura 3-10. En concreto, la imagen se ha tomado en el video filmado en el puente. Se puede apreciar en líneas azules las divisiones, y en rojo los keypoints obtenidos por el extractor, que en este caso es SIFT.

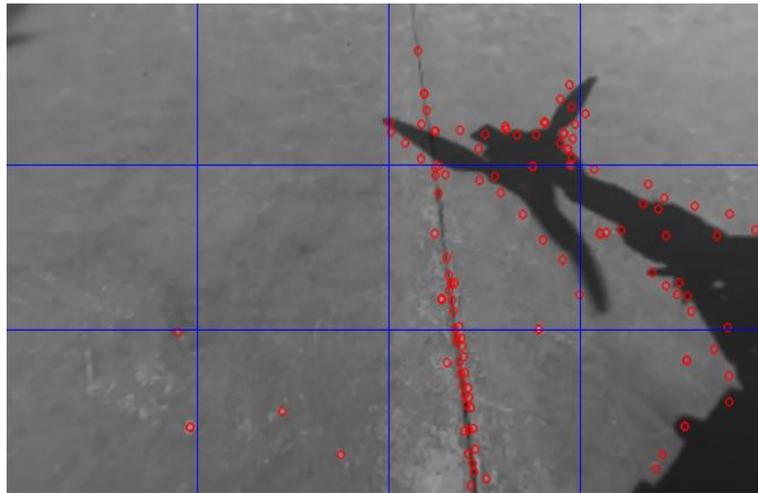


Figura 3-10 : Ejemplo división imagen en cuadrículas

Para cada uno de los cuadrados que resultan, se han contabilizado los keypoints (tantos los buenos como los malos). A partir de esto, se puede observar cuál es el cuadrante del que se obtiene más información y cuál el que menos.

3.4 Análisis de asociación de features

La asociación de features (o matching) se realiza para parejas de fotos, siendo la segunda tomada milésimas de segundos después. El emparejamiento se realiza de tal forma que la primera foto nunca se queda sin buscar su punto análogo, por tanto, la última foto que compone el video tendrá los keypoints sin enlazar.

En la veracidad de los matching juega un papel importante RANSAC, como se explicó en el apartado 2.6. Con ello, se pasa a diferenciar una asociación correcta e incorrecta. Los keypoints que logren enlazarse de manera adecuada se considerarán buenos, a diferencia de los que no lo logran.

Los análisis realizados en relación con la asociación de imágenes son:

- Número de matching totales.
- Número de matching correctos, obteniéndose además el porcentaje.
- Número de matching incorrectos con su porcentaje correspondiente.
- Matching máximos realizados en una pareja de fotos, distinguiéndose entre buenos y malos.
- Matching mínimos, diferenciándose de nuevo los buenos y no.

Todas las características anteriores se han observado para la imagen completa. Para concentrar más los resultados, se ha dividido la imagen en 12 partes iguales, tal y como se muestra en la figura 3-10. En cada uno de los cuadros resultantes se ha calculado los matching que se realizan, contabilizándose aquellos que se realicen en el mismo cuadrante. De nuevo, se diferencia entre buenos y malos. Aquellos matching que se realicen entre cuadrantes diferentes, también se estudian si son correctos o no.

3.5 Análisis por condiciones de contraste

Uno de los principales objetivos de este trabajo es comprobar cómo funciona la detección de características en las imágenes en función de las condiciones lumínicas. En este análisis se parte de la imagen seccionada en 12 cuadrantes, estudiándose para cada uno de los rectángulos resultantes:

- Contraste de la subimagen. Para calcular este valor, se calcula el histograma de la sección y se construye su función de distribución. Esta gráfica resultante, con valores comprendidos entre 0 y 1, se analiza en los valores 0.1 y 0.9. La diferencia de la respuesta obtenida será el valor de contraste que se estaba

buscando.

- Contraste en los cuadrantes donde se encuentran los keypoints. Una vez conocida la localización de los puntos clave de la imagen y su cuadrante correspondiente, se calcula para cada sección su valor de contraste, como se ha comentado anteriormente. Para cada contraste obtenido, se acumulan los puntos clave que se encuentren en su cuadrante correspondiente. Con esto se quiere conocer cómo funciona la localización o no de keypoints según las condiciones lumínicas del entorno.

3.6 Análisis por costes computacionales

Con este apartado se pretende principalmente ver la capacidad de cálculo de cada uno de los extractores, principalmente observándose los tiempos. Los estudios que se han realizado son:

- Tiempo de cálculo de keypoints, es decir, cuánto tiempo se ha tardado en calcular los puntos claves de una foto. Para esta característica se ha buscado tanto los tiempos máximos, mínimos y medios.
- Tiempo de cálculo de matching. Al igual que antes, se busca conocer cuánto se tarda en buscar el enlace de los keypoints, diferenciándose los tiempos máximos, mínimos y medios.
- Número de combinaciones que es capaz de obtener cada extractor, esto es, número de fotos que se consiguen emparejar. Con este dato se puede ver, conociéndose el número de imágenes de cada vídeo, cuántas imágenes es capaz de procesar cada extractor.

3.7 Conclusiones

Con todos los análisis presentados en este capítulo se pretende introducir los distintos métodos que se han seguido en este trabajo. La manera en que se han estructurado los diferentes apartados se repite en el capítulo siguiente, esta vez con los resultados obtenidos.

En la primera parte del capítulo se muestran los diferentes videos que han sido objeto de estudio, presentando algunas características relevantes como el histograma y contraste. A continuación, se ha dividido los distintos estudios según el objetivo en común. Por un lado, está el análisis de keypoints que se basa en la obtención de puntos clave. Después, se presenta el análisis de asociación de features, cuyo interés está basado en la relación de imágenes. Le sigue el estudio basado en las condiciones lumínicas, en concreto se estudia el contraste de cada foto. Por último, se analiza la respuesta temporal y los costes computacionales de los extractores.

En el capítulo siguiente se pasa a mostrar tablas y gráficas de todos los análisis que aquí se han presentado. Las conclusiones obtenidas se muestran en el capítulo 5, donde se podrá comprobar cuál extractor es mejor según el contraste, cuál obtiene mejores resultados, cuál es más rápido etc.

4 DISCUSIÓN

4.1 Introducción

En este capítulo se muestran los resultados obtenidos tras someter a los diferentes videos a las pruebas presentadas en el capítulo anterior. El objetivo es comprobar el comportamiento de los extractores visuales (SURF, SIFT y ORB) en diferentes escenarios. Para ello, se disponen de tres videos con diferentes características principalmente lumínicas, estudiándose para cada uno de los extractores. Los resultados obtenidos se han trabajado con Excel, obteniéndose diferentes gráficas.

La estructura de este capítulo es igual al capítulo 3. En primer lugar, se analiza el número de keypoints seguido de la asociación de imágenes. Se continúa analizando las condiciones de contraste de cada video, finalizándose el capítulo estudiando el coste computacional de cada uno de los extractores, lo que se resume en observar los tiempos de cálculo de todo lo anterior.

Cada gráfica se ha dividido en tres, correspondiente a cada vídeo. Para cada sección se muestra el resultado de cada extractor, que se define por un color mostrado en la leyenda de la gráfica.

4.2 Resultados del análisis del número de keypoints

En este apartado se muestran las gráficas obtenidas tras analizar los keypoints de cada vídeo; el número total, las distancias, la capacidad de enlace etc. Lo interesante es encontrar el extractor que sea capaz de encontrar buenos keypoints, es decir, que tengan una alta probabilidad de enlazarse de manera correcta con otra imagen. Se pasa a continuación a mostrar las diferentes gráficas obtenidas.

1. Número total de keypoints

En la figura 4-1 se muestra para los diferentes vídeos y extractores, el número total de puntos clave obtenidos.

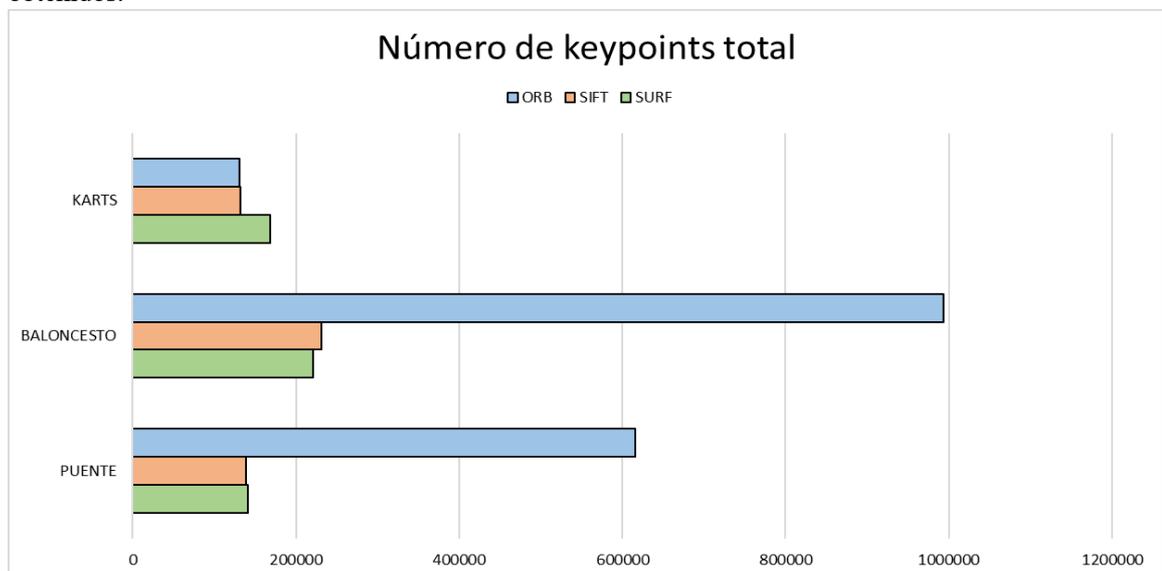


Figura 4-1 : Gráfica número de keypoints total

Como se puede observar, el extractor que obtiene más información de los vídeos es ORB, seguido de SIFT y SURF, excepto en el vídeo titulado como karts, cuyo primer puesto está ocupado por SURF. A pesar de la gráfica mostrada se debe tener en cuenta que estos resultados muestran datos repetidos, tal y como se explica en la página 6; los extractores SIFT y ORB obtienen puntos repetidos con diferentes orientaciones.

Se muestra a continuación en la figura 4-2 el caso de repetición de puntos.

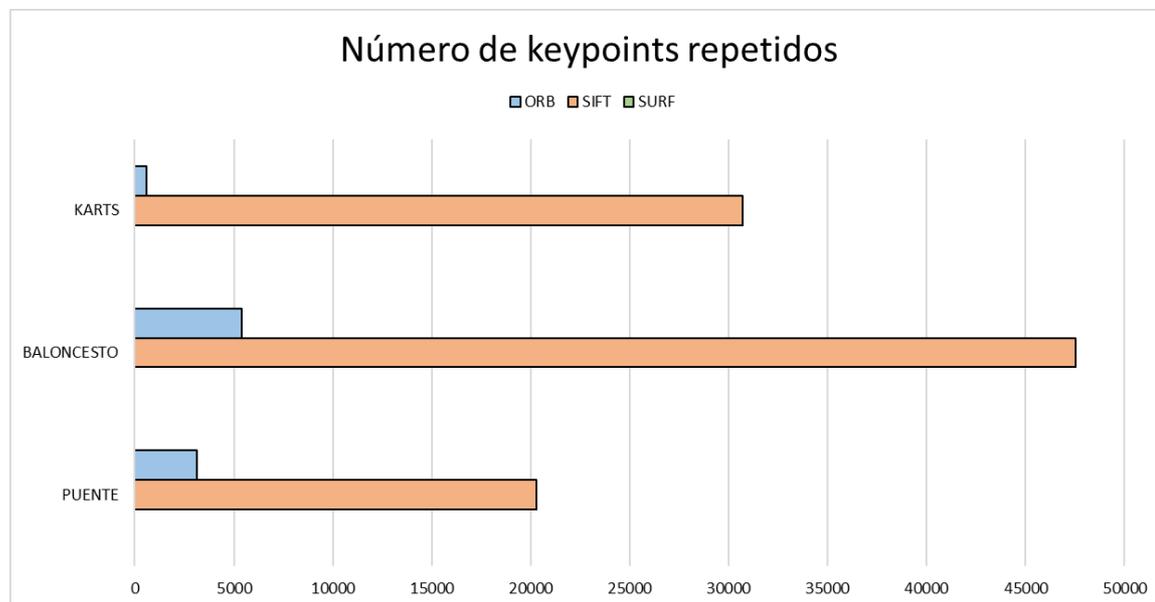


Figura 4-2 : Gráfica de keypoints repetidos

Se puede observar que en SURF no se obtienen repetición de puntos. Además, en ORB apenas ocurre esta situación comparado con SIFT. Se muestra a continuación en la tabla 4-1 el porcentaje de repetición de puntos, obtenidos como la división del número de puntos repetidos respecto al total.

	Puente (%)	Baloncesto (%)	Karts (%)
SIFT	14.59	20.68	23.31
SURF	0	0	0
ORB	0.51	0.54	0.47

Tabla 4-1 : Porcentaje de repetición de keypoints

Como se puede observar, el caso de repetición de puntos en el extractor ORB apenas es apreciable. Lo interesante es conocer los datos reales resultado de restar las dos gráficas (4-1 y 4-2). Es por ello por lo que se expone en la figura 4-3 dicho resultado.

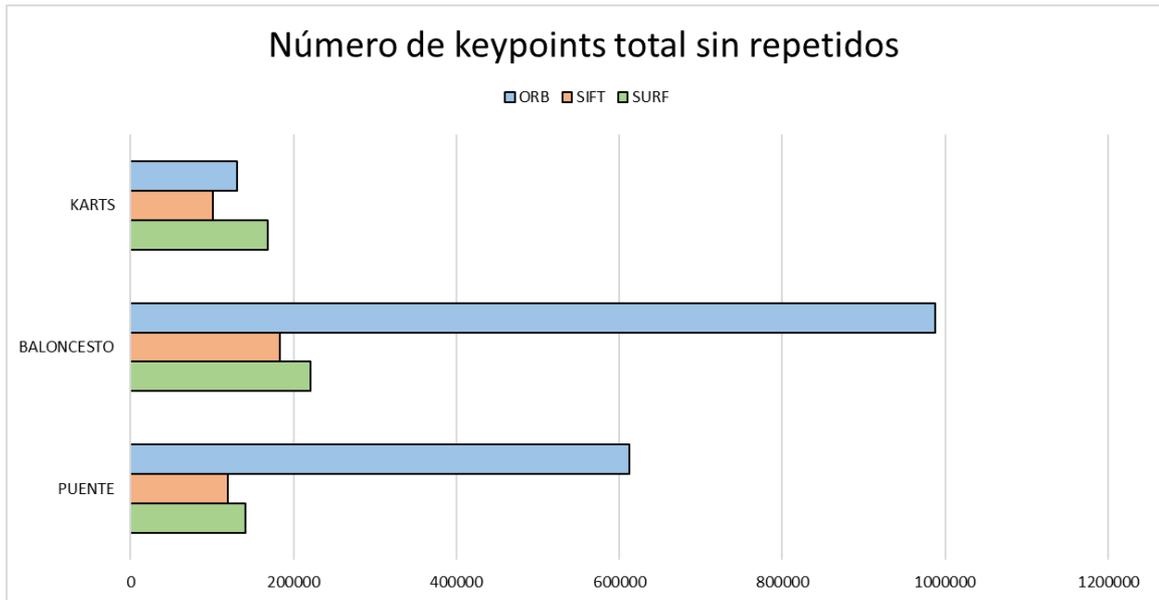


Figura 4-3 : Gráfica número total de keypoints, descartando repetidos

De nuevo se vuelve a observar que el extractor que obtiene más puntos es ORB. A pesar de ello, en el video karts se repite la situación que se comentaba en la figura 4-1, con el extractor SURF en primer lugar.

2. Número de keypoints buenos

Se considera un keypoint correcto aquel que es capaz de hacer un match bueno, es decir que encuentra de manera acertada a su pareja en la otra foto. Para saber si es correcto o no se ha usado en la programación RANSAC. En la figura 4-4 se muestra la gráfica obtenida tras este estudio.

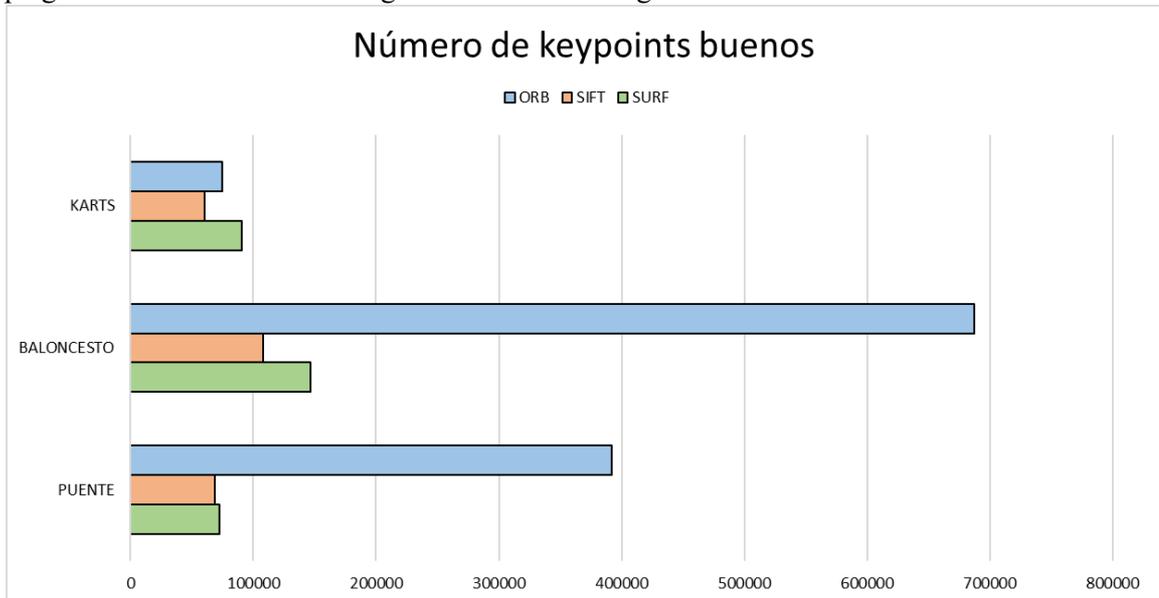


Figura 4-4 : Gráfica keypoints buenos

Se observa que en todos los videos, excepto el perteneciente a la pista de karts, el descriptor ORB ocupa en primer lugar. Como ocurría en la figura 4-1, el extractor SURF obtiene más puntos buenos en comparación con los demás. De nuevo, hay que tener en cuenta que existen puntos repetidos. Debido a que no es concebible que un punto haga match bueno más de una ocasión, a los puntos repetidos se les

pasa a llamar falsos positivos. Esta situación se pasa a mostrar en la figura 4-5.

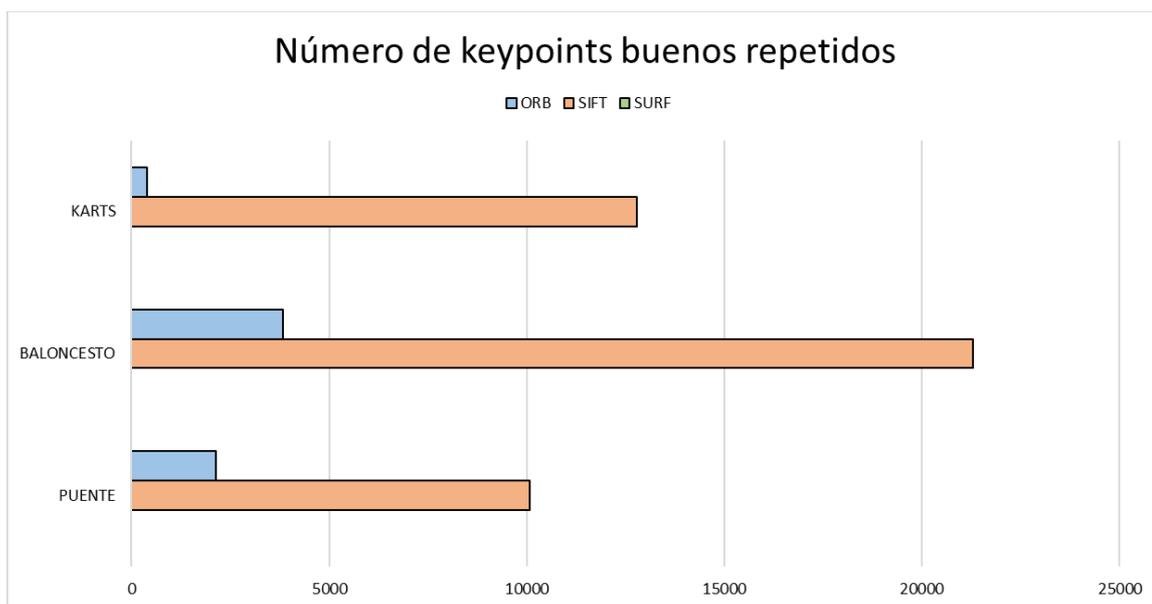


Figura 4-5 : Gráfica de keypoints buenos repetidos

Como se ha comentado anteriormente, esta situación también se la conoce como falso positivo ya que un punto no puede enlazarse con otro y resultar también un match correcto. Se busca tener un extractor con índice porcentual de falsos positivos bajo. Se muestran dichos porcentajes en la tabla 4-2, calculados como la división de los falsos positivos entre el total de puntos buenos encontrados por cada extractor.

	Puente (%)	Baloncesto (%)	Karts (%)
SIFT	14.63	19.72	21.29
SURF	0	0	0
ORB	0.55	0.56	0.52

Tabla 4-2 : Porcentaje de falsos positivos en los puntos buenos

Se pasa a exponer la gráfica 4-6, resultado de quitar los puntos repetidos, tal y como se mostró en la figura 4-3; en este caso, se muestra la diferencia entre las gráficas 4-4 y 4-5.

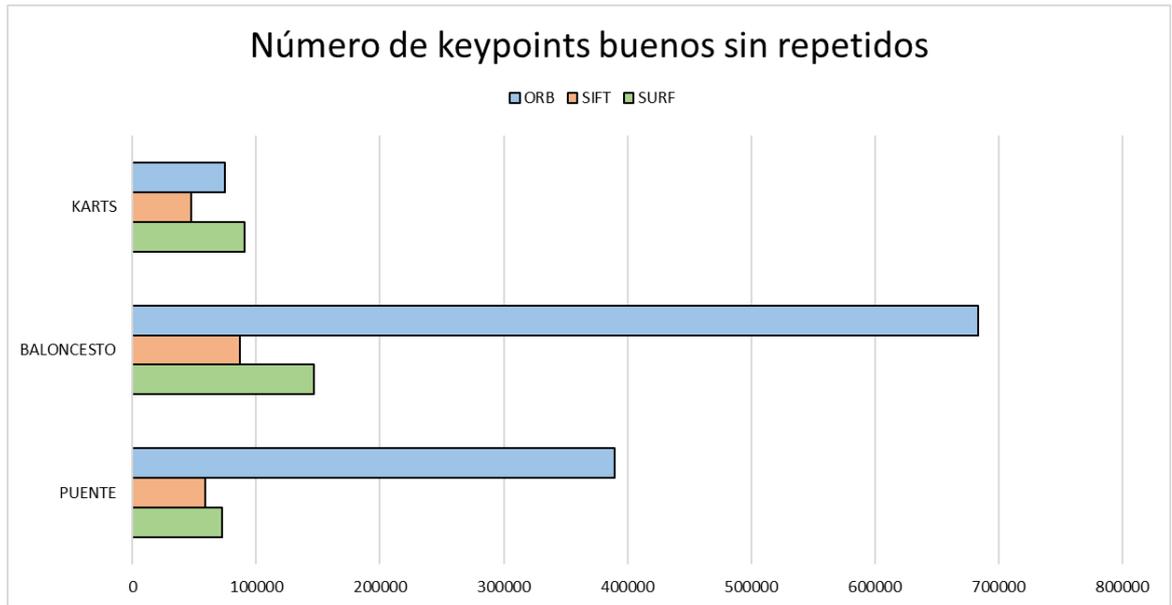


Figura 4-6 : Gráfica de keypoints buenos sin repetidos

Por último, se pasa a poner un resumen en la tabla 4-3 del porcentaje de keypoints buenos calculados para cada extractor, quitando los falsos positivos.

	Puente (%)	Baloncesto (%)	Karts (%)
SIFT	49.45	47.30	46.85
SURF	51.26	66.31	53.64
ORB	63.62	69.18	56.99

Tabla 4-3 : Porcentaje de keypoints buenos sin repetidos

Se observa que, a pesar de la repetición de puntos y obtener falsos positivos, el extractor ORB alcanza el mayor porcentaje de keypoints buenos, seguido de SURF. El extractor SIFT, que fue el primero que se desarrolló, al igual que le ocurre a ORB, obtiene repetición de coordenadas, pero los resultados que se observan en la tabla anterior lo posicionan en la última posición.

3. Número de keypoints malos

Se van a mostrar a continuación las gráficas contrarias anteriores; es decir, se obtienen tras restar al número total de keypoints los correspondientes que tengan la característica de bueno. En primer lugar, se muestra en la figura 4-7 el número total de puntos malos.

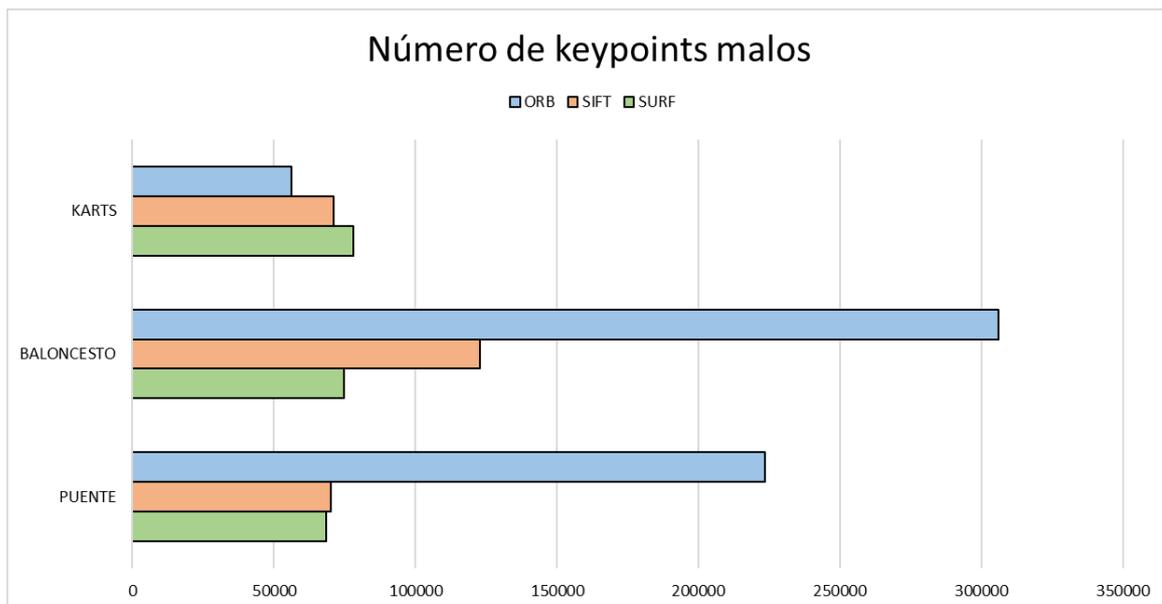


Figura 4-7 : Gráfica número de keypoints malos

El extractor ORB obtiene de nuevo en la mayoría de los videos un gran número de keypoints malos. Debido a que los extractores SIFT y ORB pueden encontrar puntos repetidos, se muestra en la siguiente figura (4-8) los resultados obtenidos. En este caso, no se les llama falsos positivos ya que puede suceder que un punto se repita y también sea un punto incorrecto.

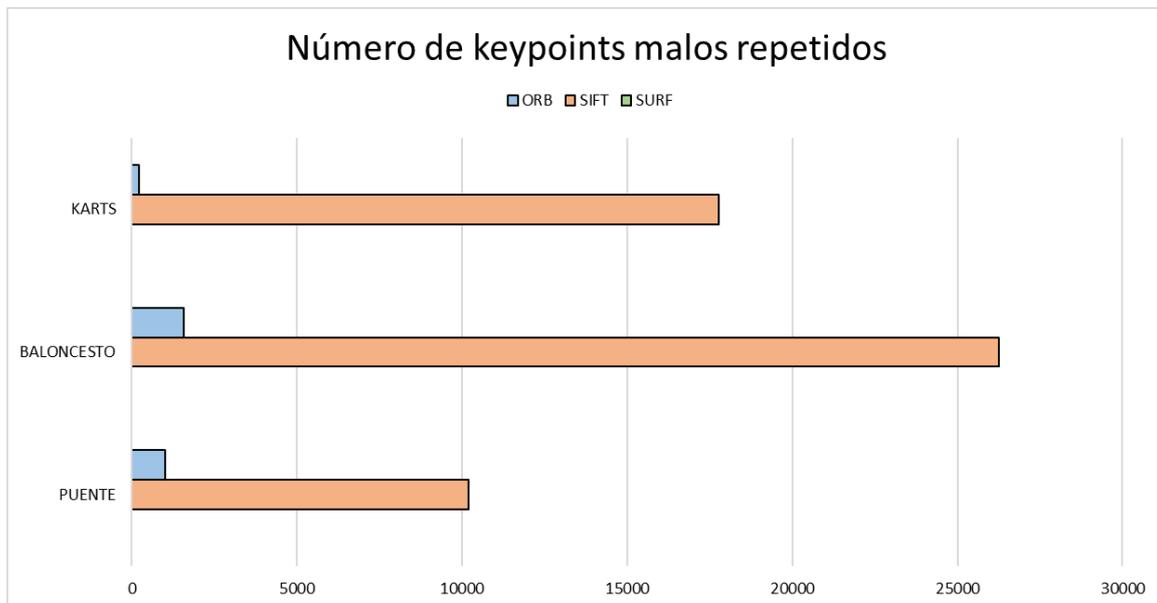


Figura 4-8 : Gráfica keypoints malos repetidos

En esta ocasión, se vuelve a observar que en SIFT se obtiene un gran número de repetición de puntos. Para tener una mejor visión de los datos, se muestra en la tabla 4-4 los porcentajes obtenidos de puntos incorrectos repetidos en comparación con el total de puntos malos.

	Puente (%)	Baloncesto (%)	Karts (%)
SIFT	14.55	21.38	24.97
SURF	0	0	0
ORB	0.45	0.51	0.41

Tabla 4-4 : Porcentaje de keypoints malos repetidos

Como ocurrió en el estudio de puntos buenos, el extractor que obtiene mayor repetición de puntos es SIFT, observándose además que ORB apenas llega al 1%. Se expone a continuación en la figura 4-9 el resultado de restar los puntos repetidos malos del total.

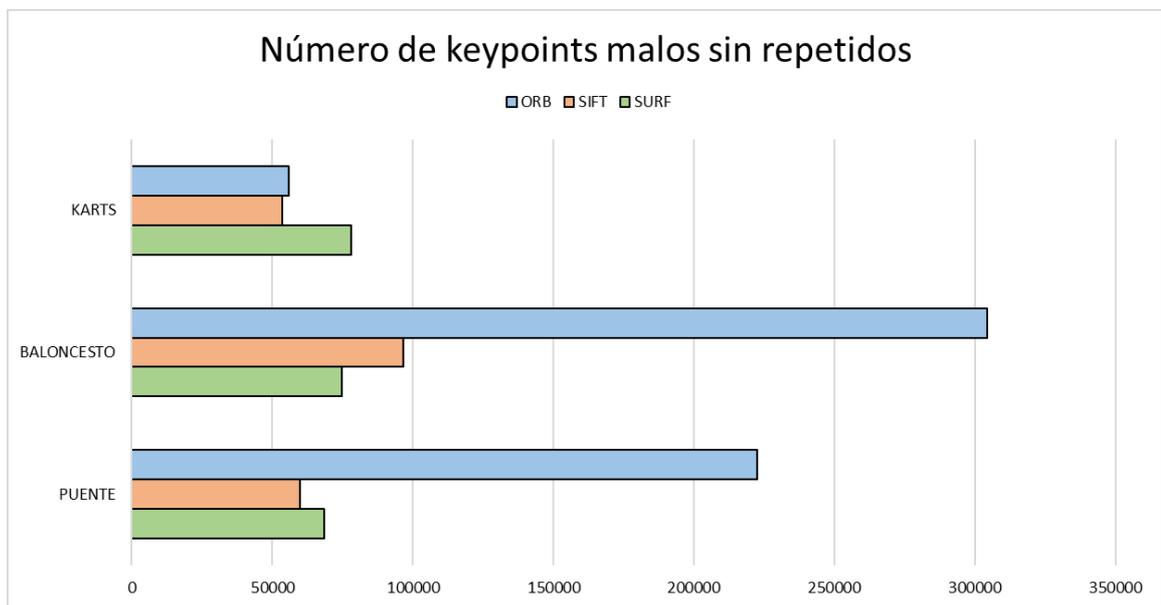


Figura 4-9 : Gráfica número de keypoints malos sin repetidos

El extractor ORB resulta ser el que más puntos malos calcula, excepto en el video karts. Lo relevante es conocer el porcentaje de keypoints incorrectos que se obtienen con relación al total de puntos obtenidos por cada extractor, tal y como se ve en la tabla 4-5.

	Puente (%)	Baloncesto (%)	Karts (%)
SIFT	50.45	52.69	52.88
SURF	48.65	33.69	46.23
ORB	36.34	30.81	42.71

Tabla 4-5 : Porcentaje de keypoints malos sin repetidos

El objetivo es tener un extractor con pocos keypoints malos, que en este caso ocurre para ORB. La suma de las tablas 4-5 y 4-3 debería dar el 100% de puntos. Este hecho no ocurre ya que para la última foto del vídeo no se pudo calcular la característica del punto, es decir, si es bueno o es malo. Esto es debido

a que siempre que se ejecute el programa, cada foto se le va a buscar una siguiente, caso que no ocurre con la última. Para que la suma de las tablas sea lo esperado, se debería tener en cuenta el número de puntos exceptuando los correspondientes a la última foto.

4. Número máximo de keypoints

En este subapartado se pasa a calcular el máximo número de keypoints que es capaz de calcular cada extractor. Además, para ese caso se diferencia entre buenos y malos. Otro máximo que se calcula es el de puntos buenos y malos.

En primer lugar, se muestra en la figura 4-10 el número máximo global que calcula cada extractor. Para el caso de SIFT y ORB se han restado los repetidos.

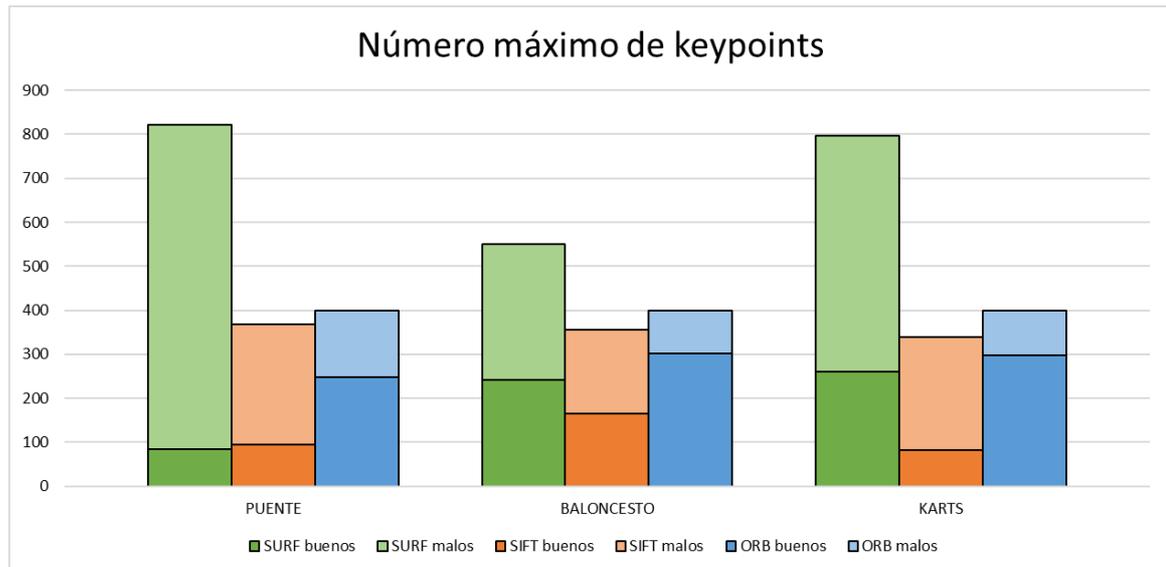


Figura 4-10 : Gráfica número máximo de keypoints con respectivos buenos y malos

En la gráfica anterior se puede apreciar que el extractor que obtiene un máximo de puntos clave es SURF, con gran parte de ellos de carácter “punto malo”. Se observa además que ORB se queda siempre clavado en el punto 400, seguido de SIFT que ronda dicho valor. Esto es debido al funcionamiento del programa, en concreto, al crear los diferentes detectores. En dicha creación, se usa una variable denominada hessiana, que se establece en el valor 400; este valor se ha elegido así ya que es un valor estándar. Cuanto más grande sea este umbral, se obtienen menos puntos clave y mayor repetición de ellos. Cada extractor funciona diferente, por eso algunos sobrepasan dicho número o lo usan como límite.

Se pasa ahora a mostrar el número máximo de keypoints buenos que cada uno de los extractores es capaz de calcular, mostrándose el resultado en la figura 4-11.

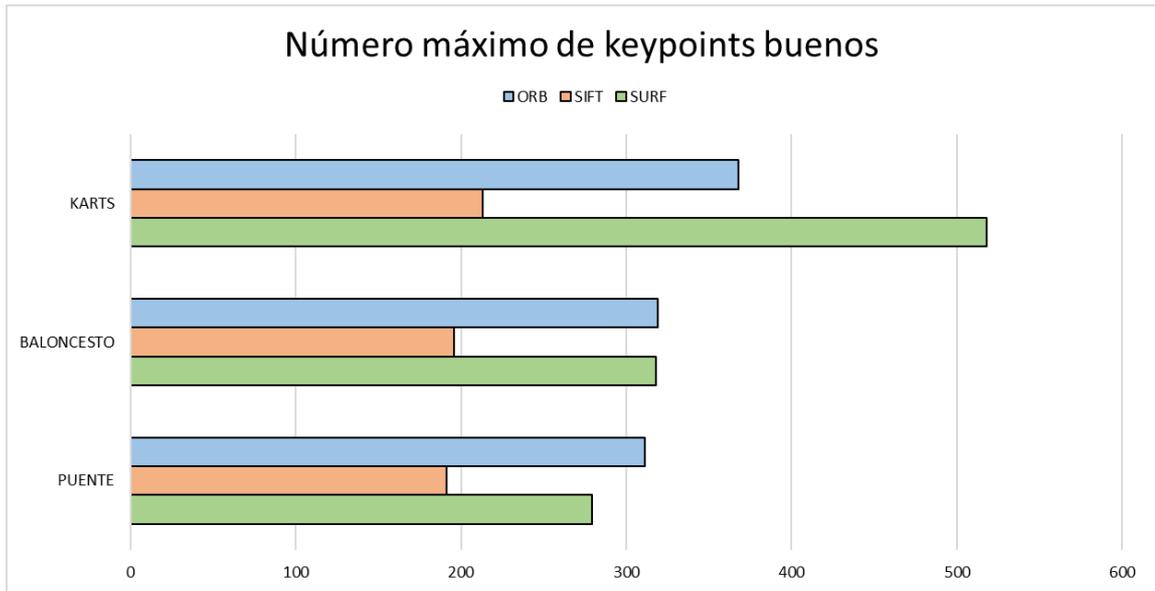


Figura 4-11 : Gráfica número máximo de keypoints buenos

Se aprecia que SIFT obtiene el menor máximo de keypoints buenos, a diferencia de SURF. El extractor ORB está prácticamente empatado con SURF, excepto en el vídeo grabado en la pista de karts, donde se observa una gran diferencia

Por último, también se muestra en este subapartado el número máximo de keypoints malos, tal y como se aprecia en la figura 4-12.

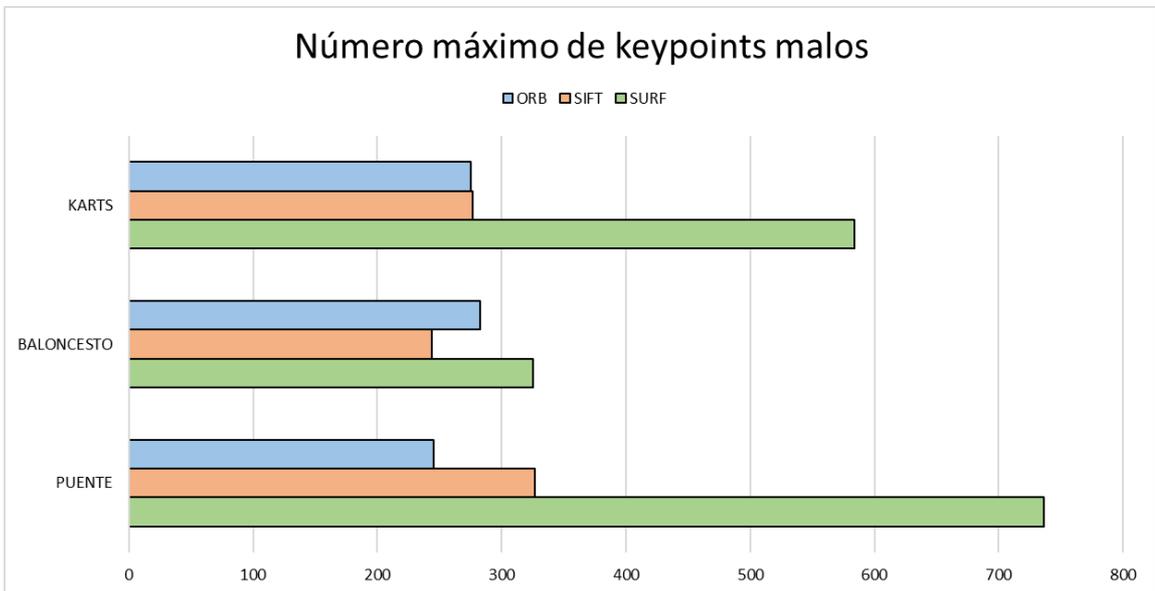


Figura 4-12 : Gráfica número máximo de keypoints malos

Como en las gráficas anteriores de este subapartado, SURF obtiene el mayor número de keypoints malos. Lo interesante es obtener un extractor que obtenga pocos keypoints incorrectos, atribuyéndose esta característica a ORB, que obtiene respuestas similares a SIFT aunque con una diferencia notoria en el video del puente.

5. Número mínimo de keypoints

Al igual que el apartado anterior, se pretende en este subapartado calcular el número mínimo de puntos que se puede contabilizar en los videos. Se diferencia además entre buenos y malos correspondientes al mínimo encontrado. Otro mínimo que se calcula será el de buenos y malos globales que se puede calcular en la duración de los videos.

En la figura que se muestra a continuación (4-13) se muestra el número mínimo de puntos que es capaz de calcular cada uno de los extractores en los diferentes videos. Dicho valor está dividido en dos, los keypoints buenos y los malos.

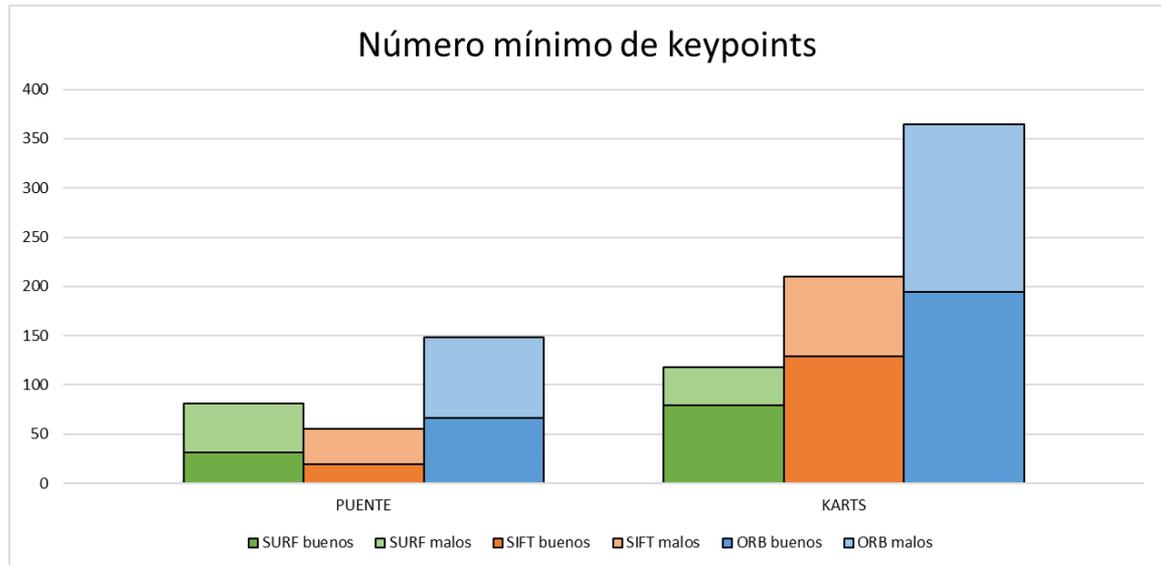


Figura 4-13 : Gráfica número mínimo de keypoints y correspondiente buenos y malos (puente y karts)

En el caso del estudio del mínimo global, el primer puesto lo ocupa ORB, con un número equitativo entre buenos y malos.

Debido a que los valores obtenidos para el vídeo grabado en la pista de baloncesto son muy pequeños en comparación con los otros vídeos, se ha decidido separarlos de la gráfica anterior para así mostrarlos en la figura 4-14. Esto es debido a que este vídeo se caracteriza por tener algunos planos dirigidos al suelo, apreciándose un solo color en la imagen y por tanto no se destacan puntos significativos.

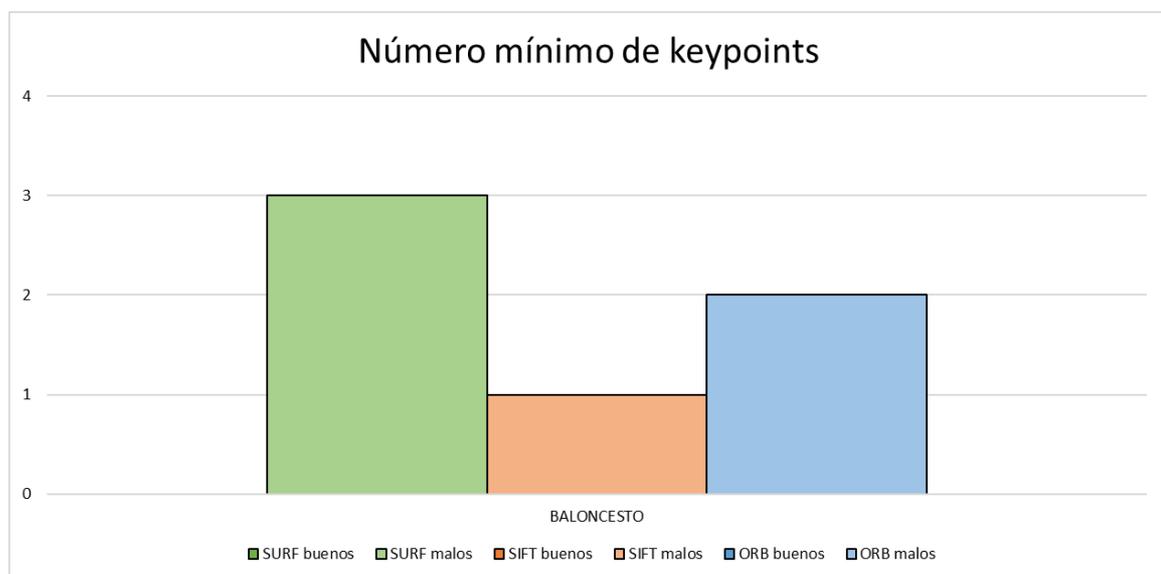


Figura 4-14 : Gráfica número mínimo de keypoints y correspondiente buenos y malos (baloncesto)

Como se acaba de comentar, en este vídeo existen algunas fotos que son monocromáticas, lo que lleva a no tener apenas keypoints. Además, como se observa en la gráfica, los pocos puntos que se encuentran son todos malos. El extractor que obtiene mayor mínimo de keypoints vuelve a ser SURF, y el que obtiene el último lugar es SIFT.

Otro factor que se ha calculado ha sido el número mínimo de keypoints buenos que se pueden encontrar en la duración de cada vídeo, tal y como se muestra en la figura 4-15.

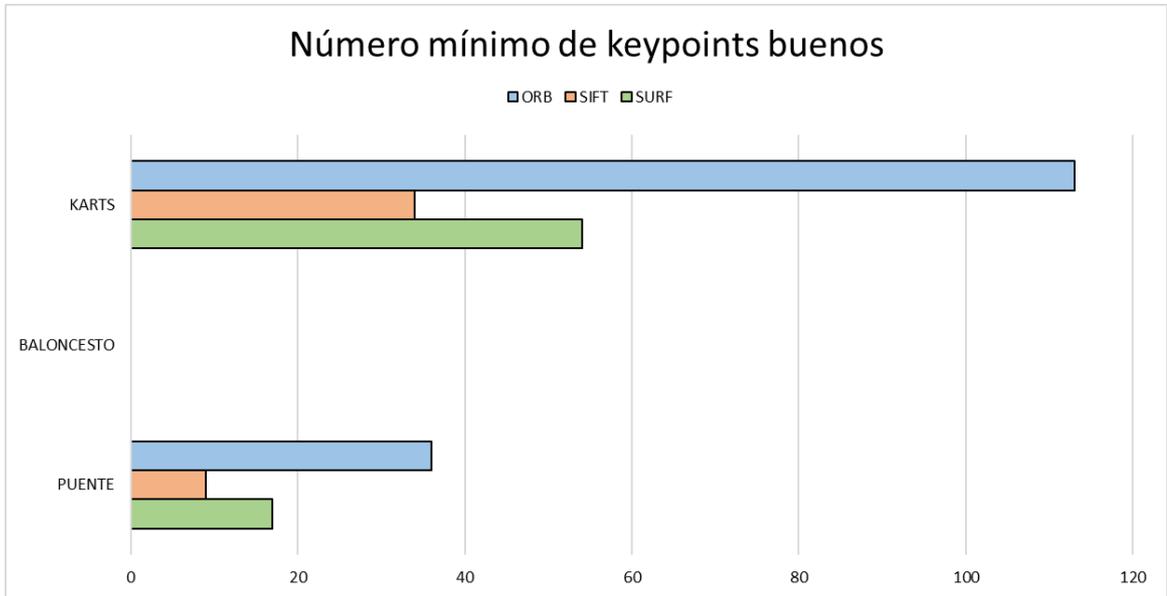


Figura 4-15 : Gráfica número mínimo de keypoints malos

De nuevo, en el vídeo grabado en la cancha de baloncesto, se aprecia que el número mínimo es cero; como ya se ha comentado, esto es debido a la existencia de fotos sin información. El extractor ORB resulta ser el que mayor mínimo de puntos correctos calcula, seguido a lo lejos por SURF.

Para finalizar, se muestra también en la figura 4-16 el número mínimo de puntos malos.

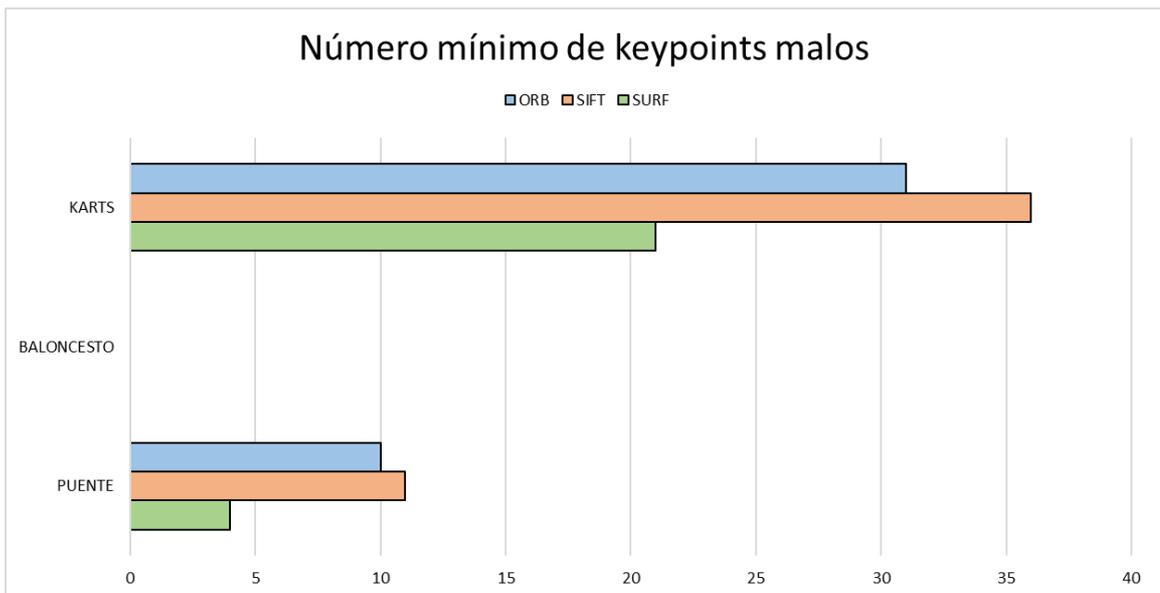


Figura 4-16 : Gráfica número mínimo de keypoints malos

En esta gráfica se observa que la situación en la cancha de baloncesto se repite y que, en este caso, el extractor SIFT obtiene un mínimo mayor en comparación con el resto, seguido de ORB y de SURF. Se

persigue que el extractor más adecuado debería tener pocos puntos malos, lo que se traduce en tener un mínimo de puntos incorrectos bajo, lo que ocurre con SURF. Para los puntos buenos, se busca lo contrario, es decir, un mínimo alto, que como se aprecia en la figura 4-15 ocurre con ORB.

6. Distancia máxima entre keypoints

Una vez que se conocen las coordenadas de los puntos, se ha calculado la distancia entre ellos. En los subapartados que siguen se pasará a calcular la distancia máxima, mínima y media. En concreto, en este subapartado se mide la distancia máxima. Se pretende conocer si los puntos están muy espaciados en la imagen, aunque este aspecto se observa mejor en el subapartado de distancias medias. Se pasa a mostrar en la figura 4-17 la distancia máxima entre keypoints.

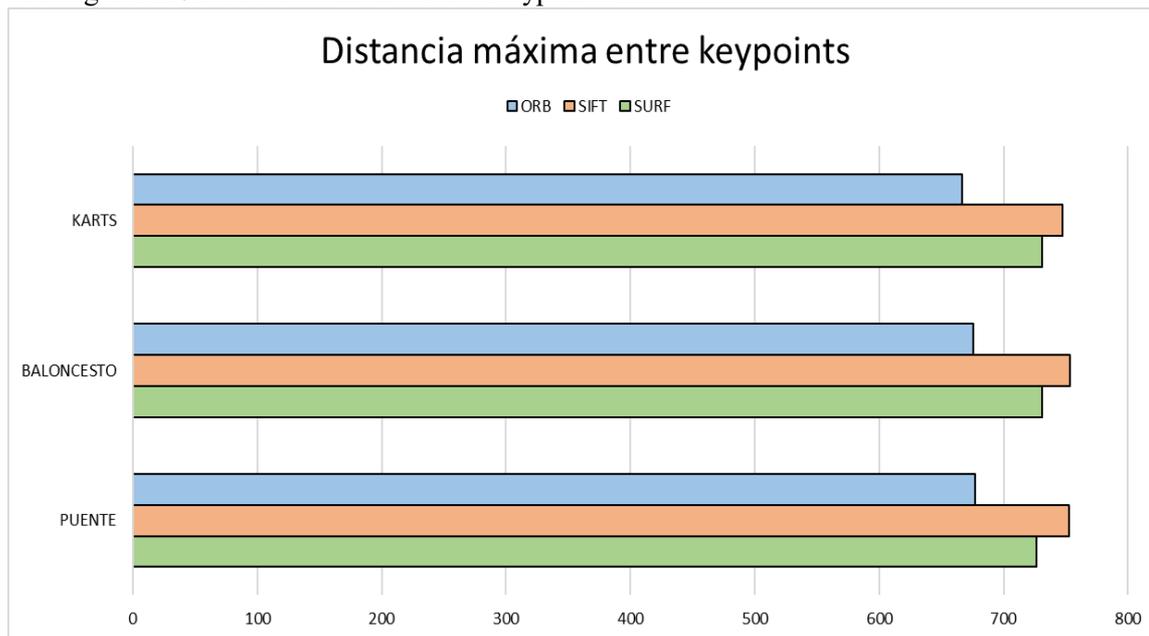


Figura 4-17 : Gráfica distancia máxima entre keypoints

Se puede observar que con el extractor SIFT se consigue calcular mayores distancias en comparación con los otros extractores. Los valores obtenidos son muy similares con respecto a los relativos a la segunda posición, que en este caso es SURF. El extractor que obtiene menor distancia máxima resulta ser ORB. Además, se aprecia que para cada extractor el valor máximo que se obtiene es prácticamente igual para cada video.

7. Distancia mínima entre keypoints

En este subapartado se pasa a calcular la distancia mínima que existe entre los diferentes puntos calculados para cada foto. El resultado se muestra en la figura 4-18.

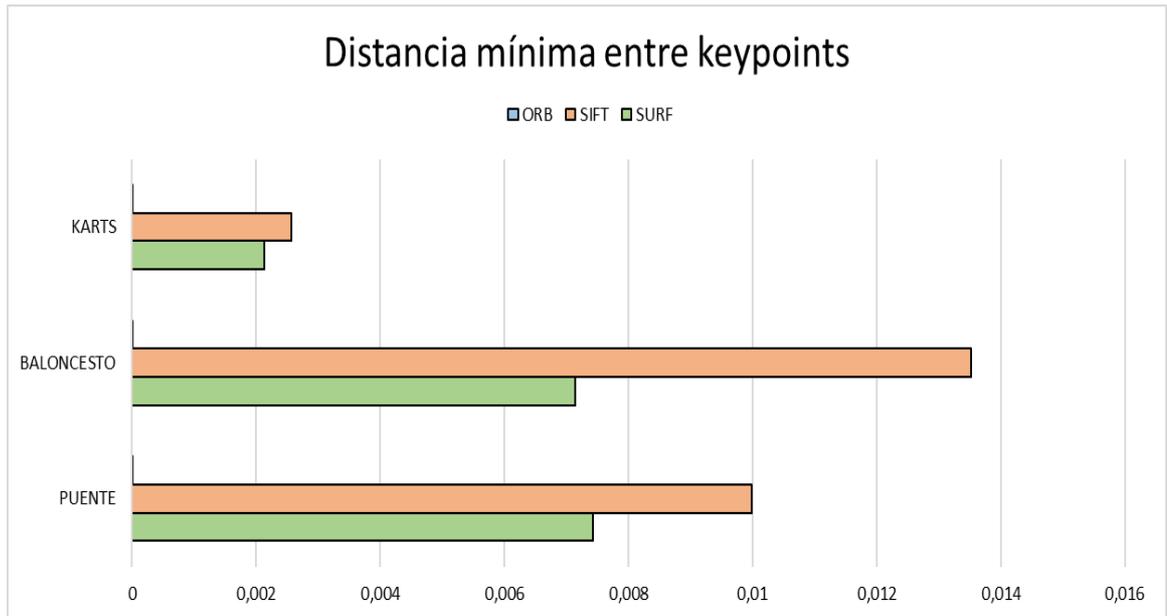


Figura 4-18 : Gráfica distancia mínima entre keypoints

Como se observa, las distancias obtenidas son muy pequeñas, con valores en torno a cero. En concreto, el extractor ORB obtiene distancias tan pequeñas que apenas se aprecia en la gráfica, lo que indica que existen puntos muy próximos a pesar de haber descartado los puntos con coordenadas iguales.

8. Distancia media

Para poder medir mejor el espacio ocupado por los puntos, se ha recurrido a calcular la distancia media entre ellos. En primer lugar, se muestra en la figura 4-19 la distancia media máxima.

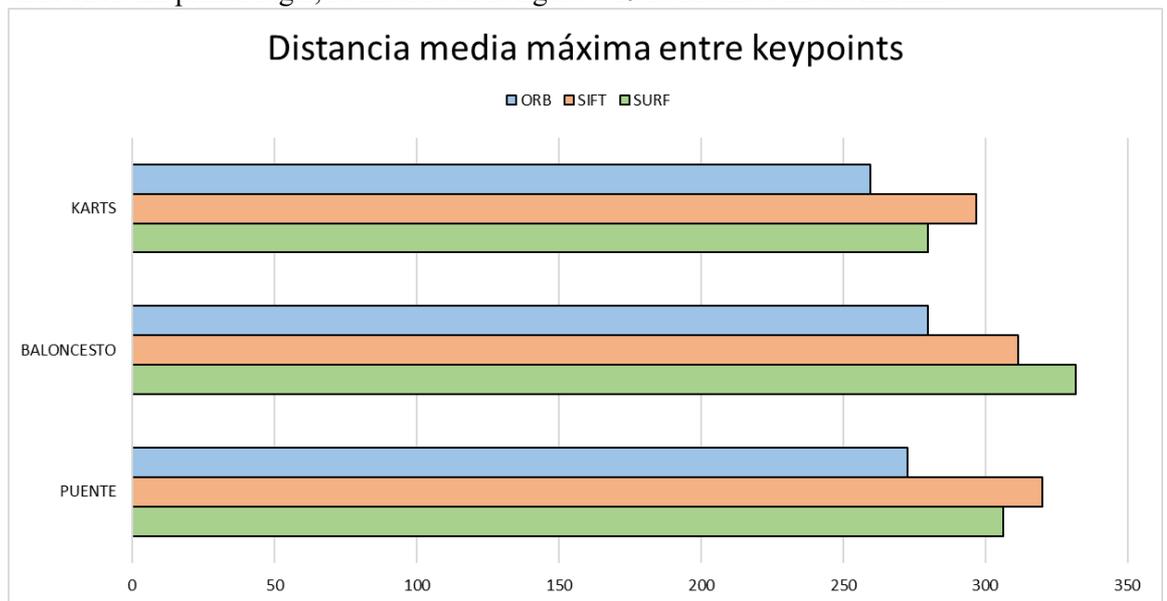


Figura 4-19 : Gráfica distancia media máxima entre keypoints

Los resultados son prácticamente iguales según el extractor, y la diferencia entre ellos no muy relevante. Con ORB se obtienen menores distancias máximas, al contrario de SIFT que obtiene las mayores excepto en el vídeo del puente.

Se pasa ahora a mostrar la distancia media mínima en la figura 4-20.

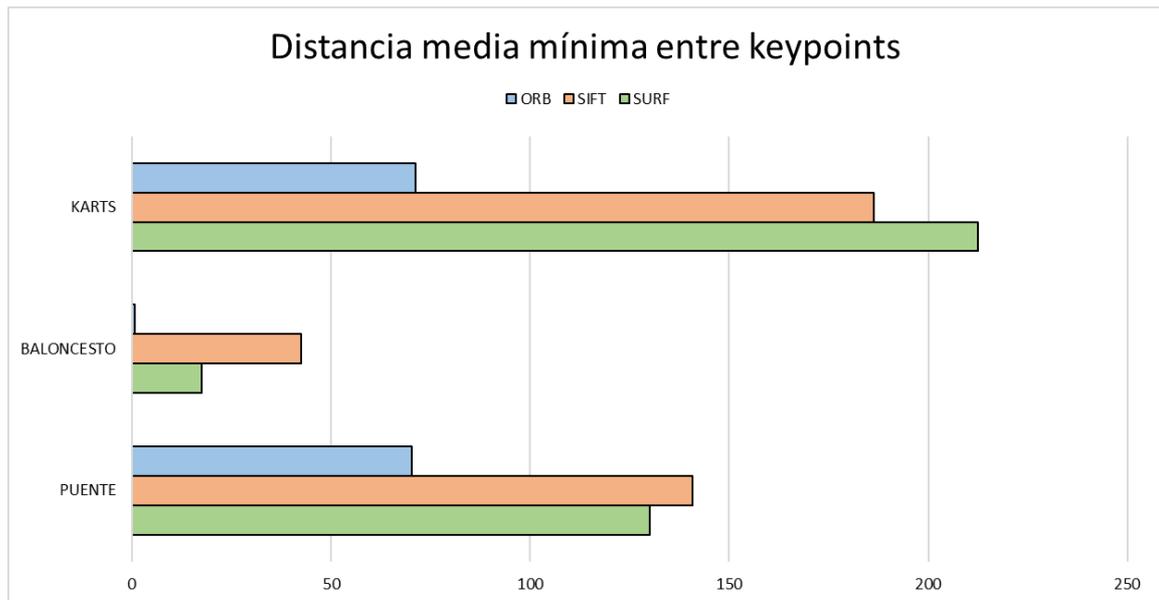


Figura 4-20 : Gráfica distancia media mínima entre keypoints

Como se puede observar, ORB obtiene mayores distancia medias mínimas, incluso en el vídeo grabado en la pista de baloncesto el valor apenas es apreciable. Esto es debido a que, observando los resultados, ocurre un momento que el extractor ha calculado sólo dos puntos, y la distancia que existe entre ellos además es muy cercana. Esta situación se observó antes en la figura 4-18, donde se aprecia que ORB obtiene unas distancias mínimas muy pequeñas.

Debido a que con las distancias medias máximas y mínimas no se puede llegar a obtener una conclusión concisa, se muestran a partir de aquí nuevas figuras en las que se recogen todas las distancias medias para cada extractor y vídeo, marcándose para cada gráfica la mediana y la media. Con la mediana lo que se busca es conocer donde se produce el cúmulo de datos, y dependiendo de la posición de la media respecto a la mediana se concluirá si se tienen los puntos muy cercanos o no.

En primer lugar, se pasa a exponer en la figura 4-21 los resultados obtenidos con SIFT en el vídeo puente.

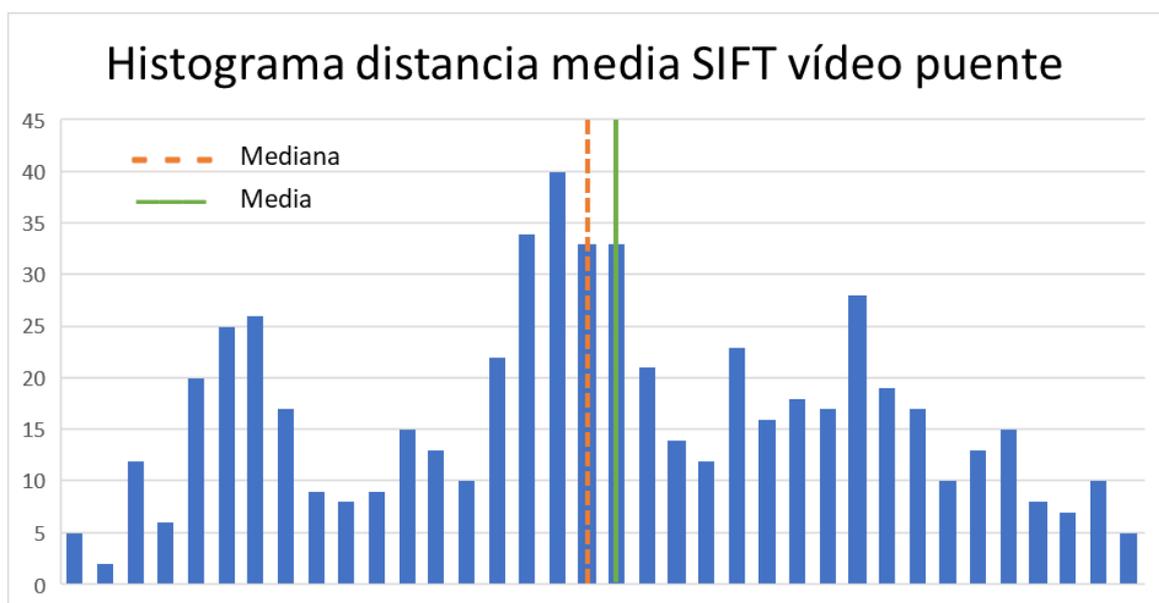


Figura 4-21 : Gráfica distancia media SIFT vídeo puente

En la figura anterior se puede observar que la mediana se marca en color naranja, mientras que a su derecha en verde se obtiene la media de todas las medias distancias calculadas. A pesar de estar muy próximas, la media se encuentra en el tramo derecho, por lo que los puntos se encuentran un poco más separados.

A continuación, en la figura 4-22 se expone los resultados de SIFT en la cancha de baloncesto.

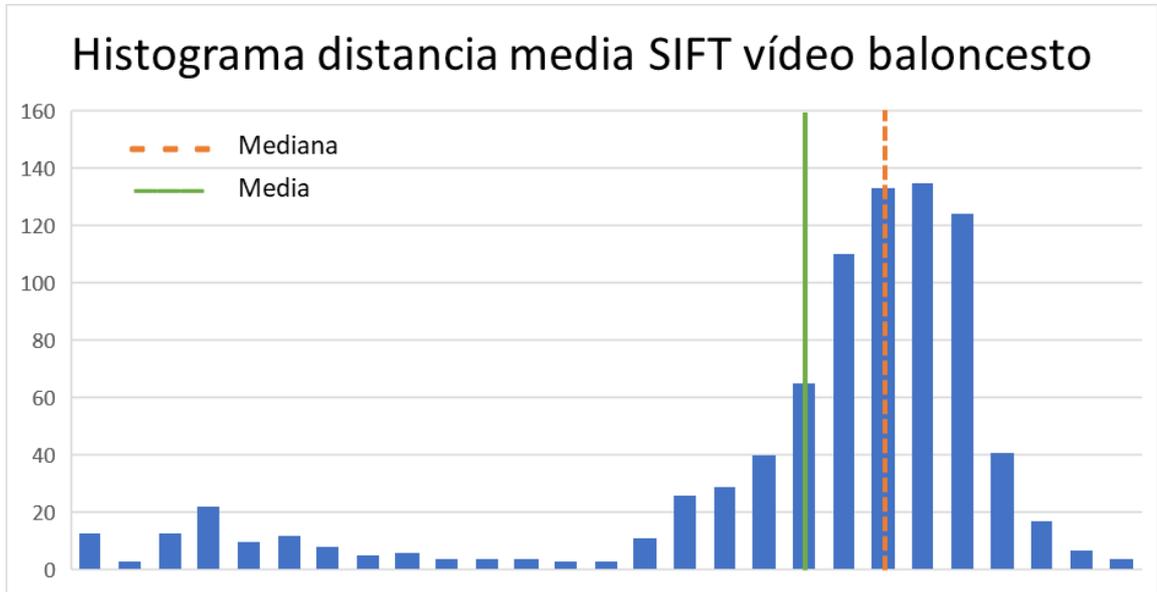


Figura 4-22 : Gráfica distancia media SIFT vídeo baloncesto

En esta grafica se observa que existe un amplio rango partiendo desde la mediana hacia la izquierda, en el que se acumula la mitad de los datos calculados. Debido a que la mediana y la media están más cerca del máximo de la gráfica, se puede concluir que las distancia entre puntos está más bien distorsionada.

Por último, para el extractor SIFT se muestra la gráfica 4-23 grabada en el vídeo de la pista de karts.

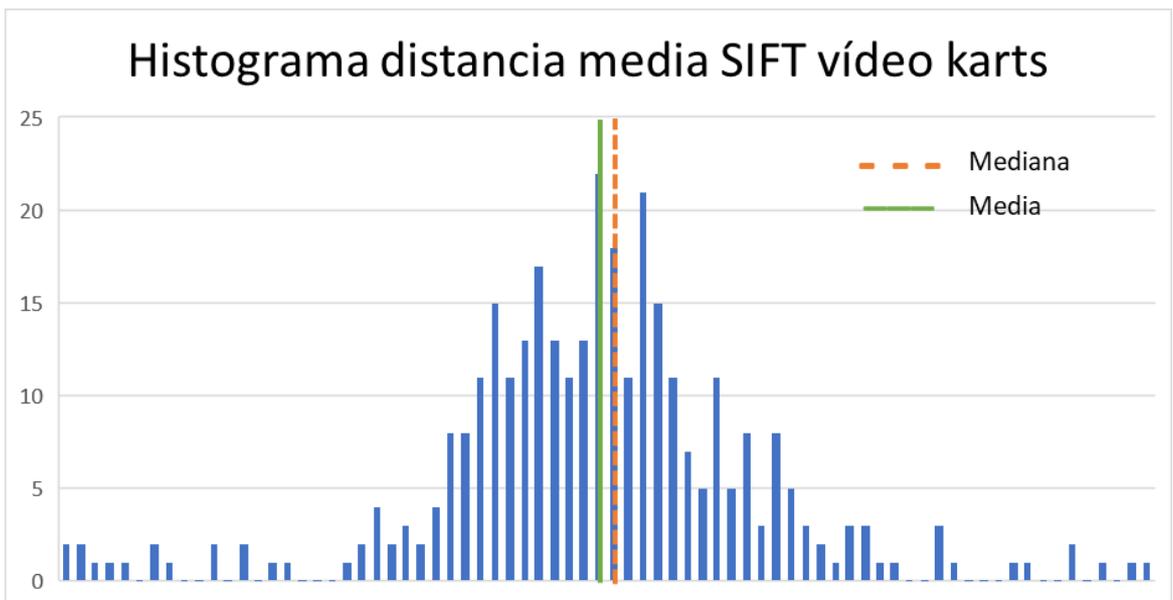


Figura 4-23 : Gráfica distancia media SIFT vídeo karts

En la figura anterior se observa que la distribución de datos obtiene una forma muy similar a la campana de Gauss, obteniéndose un cúmulo en la parte central de la gráfica. La mediana y la media están muy

próximas entre sí y en la parte central de la figura. Con toda esta información se concluye que los puntos están bien repartidos en el espacio de grabación, que no tienden a estar ni muy juntos en una zona ni muy esparcidos.

Se comienza ahora a exponer las tres gráficas correspondientes al extractor SURF, siendo la primera de ellas la número 4-24 grabada en el puente.

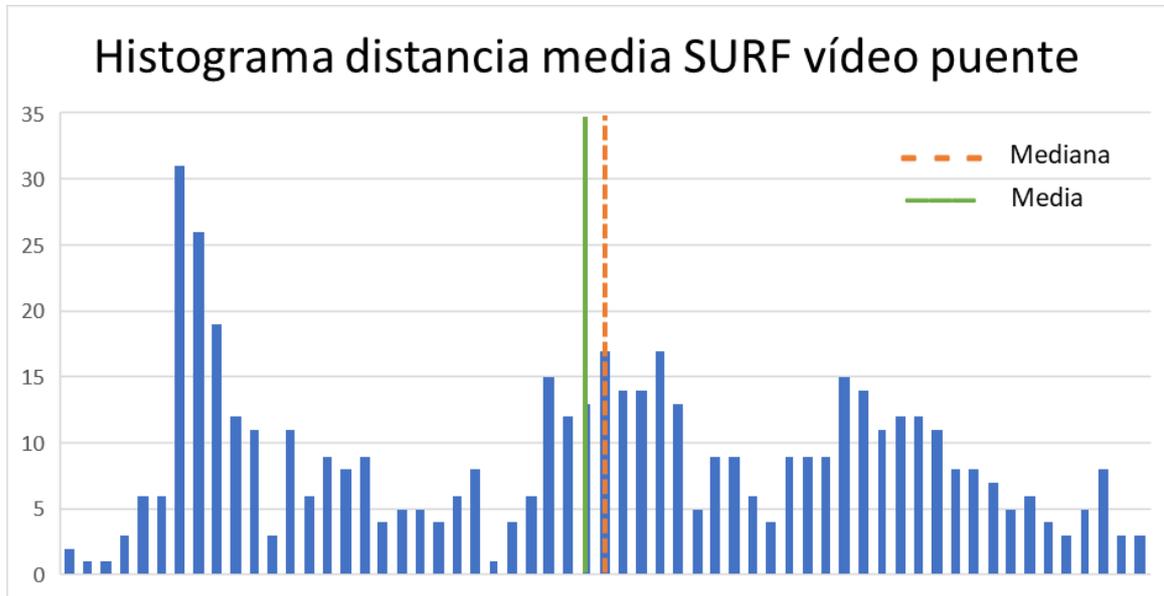


Figura 4-24 : Gráfica distancia media SURF vídeo puente

En la imagen anterior se observa que la información está prácticamente distribuida por toda la gráfica, además la media y la mediana se encuentran en la zona central. Debido a ello se concluye que los puntos se encuentran esparcidos en el espacio por toda la duración del vídeo, incluso debido a que la media está a la izquierda los puntos comienzan a juntarse un poco.

En la figura siguiente (4-25) se muestran los resultados obtenidos en la pista de baloncesto con SIFT.

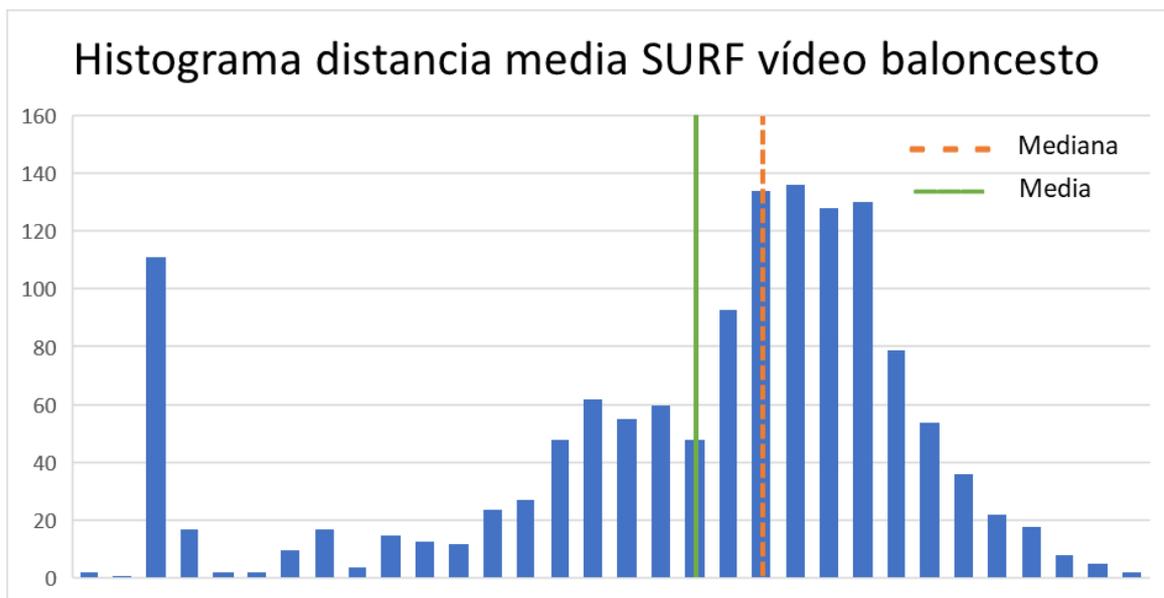


Figura 4-25 : Gráfica distancia media SURF vídeo baloncesto

En esta ocasión se aprecia que los datos desplazan el histograma a la derecha, obteniéndose más

distancias medias máximas. Esto indica que los puntos están más separados de lo normal, aunque la media de todos los datos indica que gran parte de ellos se encuentran próximos entre sí. Se puede concluir que, a pesar de que la media indique que los puntos están separados, dicha distancia comienza a ser elevada.

Por último, se ha calculado para SURF los resultados de las distancias medias en el vídeo de la pista de karting, en la imagen 4-26.

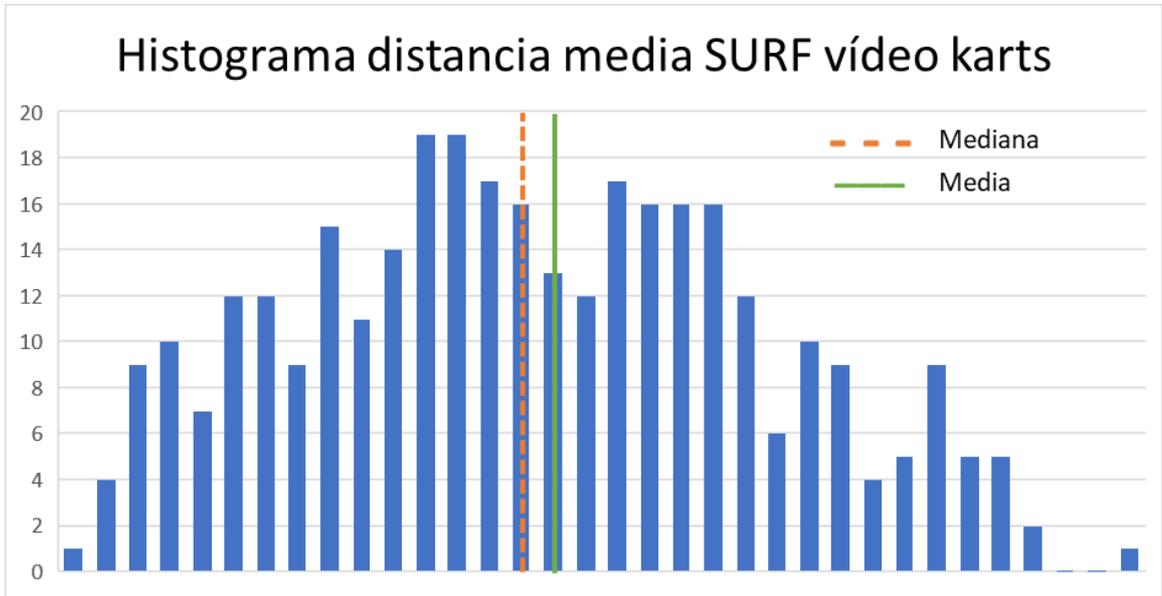


Figura 4-26 : Gráfica distancia media SURF vídeo karts

Una de las conclusiones que se llega con la figura anterior es que se tiene un gran volumen de datos, debido a que las se obtienen muchas barras y muy altas. Además, los datos se encuentran bien distribuidos a lo largo de la gráfica, con la media y la mediana marcadas en el centro de la imagen. Con ello se concluye que los puntos se encuentran próximos entre sí, que no existen zonas relevantes de cúmulos de puntos.

Se muestra a continuación una nueva tanda de 3 gráficas, esta vez pertenecen al extractor ORB. La primera de ellas (4-27) se refiere a los datos extraídos del vídeo del puente.

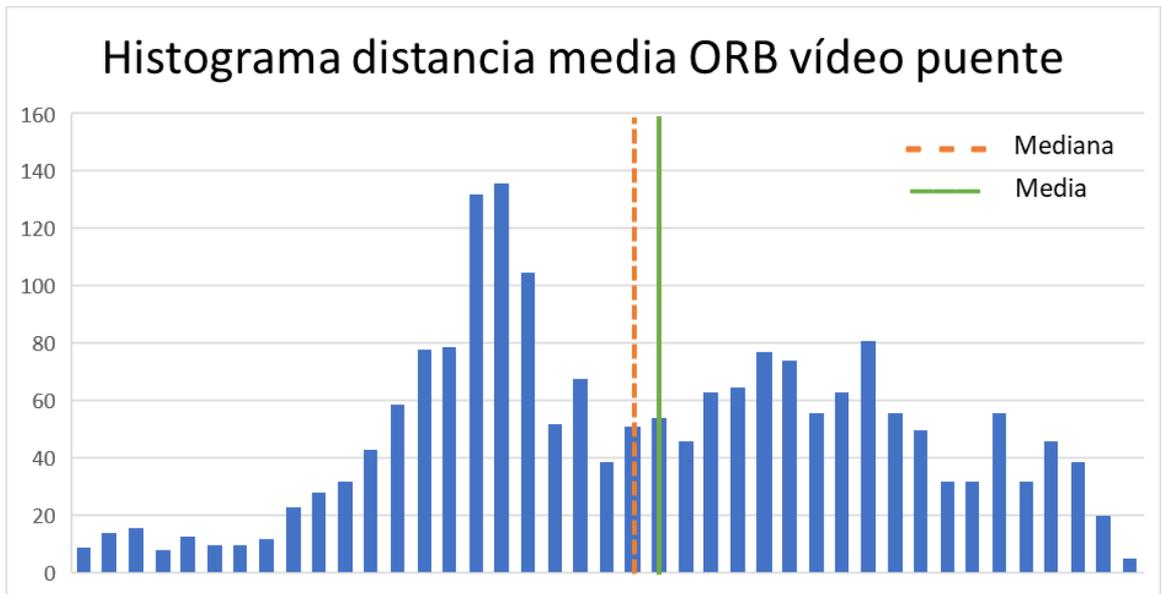


Figura 4-27 : Gráfica distancia media ORB vídeo puente

Con el extractor ORB en el vídeo del puente, se observa que los datos se encuentran distribuidos a lo largo del gráfico, aunque la media y la mediana se encuentran un poco desplazadas a la derecha. Debido a que están muy próximas entre sí y a la posición que tienen en la figura, se puede concluir que los puntos que se calculan a lo largo del vídeo se encuentran distribuidos en el espacio de grabación.

La figura siguiente (4-28) muestra los resultados obtenidos tras analizar las distancias medias que obtiene ORB en el vídeo grabado en la pista de baloncesto.

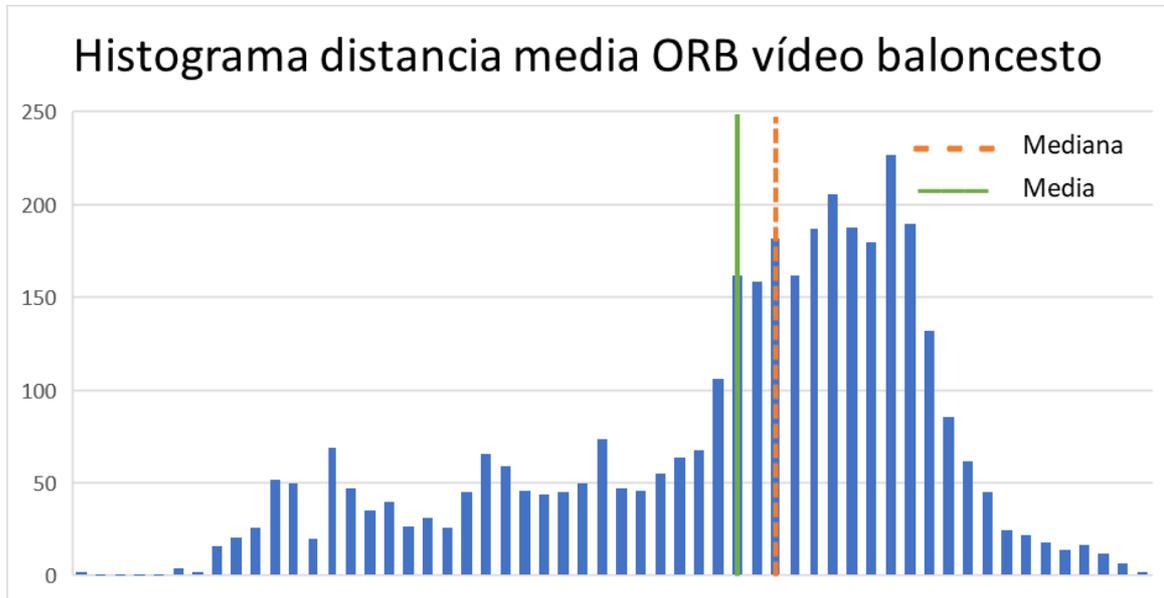


Figura 4-28 : Gráfica distancia media ORB vídeo baloncesto

En esta ocasión, se puede observar que la gráfica se encuentra desplazada hacia la derecha, indicando que las distancias que se han calculado tienden al máximo. Con ello se llega a la conclusión de que la mayoría de los puntos se encuentran muy separados entre sí.

La última figura que se muestra (4-29), recoge la información extraída del vídeo de la pista de karting para el extractor ORB.

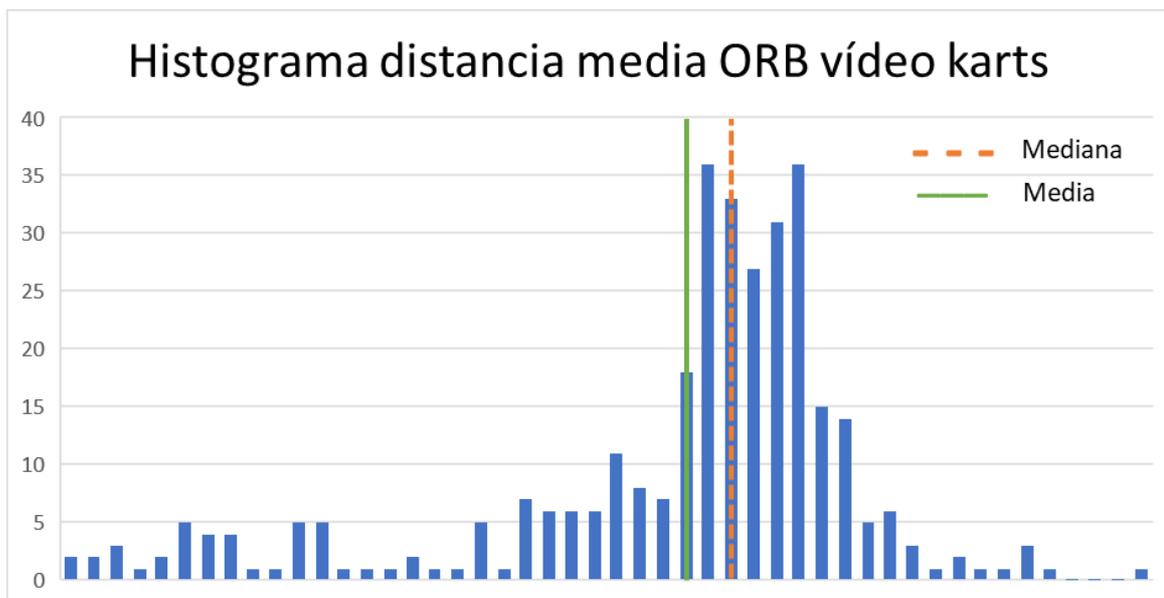


Figura 4-29 : Gráfica distancia media ORB vídeo karts

En la gráfica anterior se observa que una de las mitades del conjunto de datos abarca desde el mínimo hasta más de la mitad de la figura. Dicha información se complementa con la posición de la media y la mediana, que se encuentran cercanas al máximo, aunque a medio camino del centro de la imagen. Con todo, se puede decir que los puntos están dispersos por la imagen, aunque no en posiciones tan extremas.

9. Número de keypoints según cuadrante

Tal y como se explicó en el capítulo 3, se ha realizado otro estudio a los diferentes videos, dividiendo cada uno de ellos en 12 cuadrantes. Para cada uno de los rectángulos resultantes, se ha calculado el número de keypoints, diferenciando entre buenos y malos. Además, se ha marcado en verde el cuadrante donde ocurre el mínimo y rojo en el caso de máximo.

En primer lugar, se muestra en la figura 4-30 el resultado obtenido para el video grabado en el puente. En el caso de los extractores SURF y ORB, se han descartado las coordenadas repetidas.

SIFT			
11856	14023	15842	12676
7871	10298	9790	10350
4354	5827	7232	8658
SURF			
13703	14225	17540	13834
9693	11088	12463	12969
4341	8414	10858	11612
ORB			
37062	52192	77169	45253
45229	60875	65204	68642
22228	48622	48246	41741

Figura 4-30 : Número de keypoints total según cuadrante vídeo puente

Para este video se aprecia que todos los extractores coinciden tanto en el cuadrante con mayor número de puntos como el que menos. Además, la sección con menos keypoints resulta ser una esquina del video.

Se continúa mostrando los resultados del total de puntos clave en el vídeo de la pista de baloncesto en la ilustración 4-31.

SIFT			
22699	24750	27765	22701
10083	13681	13784	11243
9627	10379	9999	6488
SURF			
33394	32756	34651	27473
12133	15158	14872	14638
8929	10088	9727	7610
ORB			
79073	114652	147719	76680
70165	138827	122453	62287
42764	59823	46190	27659

Figura 4-31 : Número de keypoints total según cuadrante vídeo cancha baloncesto

De nuevo, todos los métodos coinciden en que el cuadrante con más y con menos puntos es el mismo. La esquina inferior derecha resulta ser en este caso la sección donde se calculan menos keypoints.

Por último, en la gráfica 4-32 se muestra los resultados para los cuadrantes del vídeo de la pista de karting.

SIFT			
9553	9792	7884	4691
6625	9084	7824	4324
7843	12368	13646	7317
SURF			
11493	13443	9819	6830
11295	14606	13191	8440
13377	23646	24413	17985
ORB			
9605	10795	8930	2699
10430	22795	20468	5784
5288	15547	13657	4662

Figura 4-32 : Número de keypoints total según cuadrante vídeo pista karts

En esta gráfica se observa los extractores no coinciden como ocurría en los vídeos anteriores. Si se observan algunas de las gráficas anteriores, por ejemplo la relacionada con contar los puntos totales (4-1), el vídeo karts obtiene respuestas no esperadas en comparación con los otros; en aquella gráfica ORB ocupaba el primer lugar excepto en el vídeo en cuestión.

A continuación, se muestran 3 nuevas gráficas en las que se recogen los datos pertenecientes a los keypoints que hacen un match correcto. La suma de las gráficas de puntos buenos y malos no dan la total, ya que en la última foto no se puede calcular si los puntos son correctos o no porque no tiene otra foto con la que enlazarse.

En primer lugar, se pasa a enseñar en la figura 4-33 el resultado de contabilizar los puntos clave correctos en los cuadrantes del vídeo rodado en el puente de la escuela.

SIFT			
5239	6477	6985	5323
4766	6193	5154	4520
2302	3495	4115	4164
SURF			
5828	6641	7925	5813
5784	6904	6732	5925
2807	5463	6508	5819
ORB			
21301	34941	45460	26850
30316	43790	40275	42591
15571	33054	31616	23860

Figura 4-33 : Número de keypoints buenos según cuadrante vídeo puente

De nuevo se aprecia que los extractores coinciden. Además, si se compara con la gráfica que recogía el total de keypoints (4-30), los cuadrantes resaltados con colores son los mismos.

En la figura 4-34 que viene a continuación, se muestran los cuadrantes rellenados para el vídeo de la pista de baloncesto, contabilizándose los puntos buenos.

SIFT			
10377	12723	13752	9934
4607	6947	7402	5444
3848	4886	4267	2472
SURF			
20985	22392	24175	18187
7939	11170	11027	9659
4671	6446	6146	4034
ORB			
48532	82594	108283	51929
46866	101695	89991	40753
26326	40224	30012	16542

Figura 4-34 : Número de keypoints buenos según cuadrante vídeo cancha baloncesto

Al igual que antes con el vídeo del puente, los extractores coinciden y además también lo hace con su gráfica de totales (4-31).

Por último, se añade también la gráfica 4-35 que marca los puntos correctos del vídeo de la pista de karting

SIFT			
4796	4868	3847	1985
2871	3760	2911	1804
3364	6486	6938	3666
SURF			
6172	7797	5397	3414
5153	6974	6031	4273
6271	14572	14268	10079
ORB			
5647	7061	5666	1413
5803	14516	13014	2817
2318	8030	6859	1313

Figura 4-35 : Número de keypoints buenos según cuadrante vídeo pista karts

Como ocurrió en la figura 4-32, ninguno de los extractores coincide en los cuadrantes y tampoco con su gráfica de totales. Las coincidencias que se tienen con la figura de totales ocurren en su totalidad con SIFT, el cuadrante donde ocurre el mínimo de SURF y el cuadrante donde ocurre el máximo de ORB. Comparando ambas gráficas, para que SURF sea igual debería estar marcado como el cuadrante de máximos el número 11, que observando los valores de la gráfica anterior es el segundo con mayor número de keypoints buenos; se concluye que la diferencia no es muy grande. Para el caso de ORB, se necesitaría que el cuadrante 4 estuviese marcado como mínimo. Como acaba de ocurrir con SURF, observando los valores se tiene que el cuadrante necesario para igualar ambas gráficas ocupa el segundo lugar de mínimo, con una diferencia con el mínimo global muy pequeña.

Para terminar este subapartado, se pasa a exponer las últimas tres gráficas, que contienen la información relevante con el número de keypoints incorrectos según vídeo y extractor. Se muestra en la figura 4-36 los resultados para el caso del vídeo del puente.

SIFT			
12322	12023	14010	12767
5476	6734	6381	5798
5779	5493	5732	4016
SURF			
12409	10361	10471	9285
4194	3988	3845	4979
4258	3642	3581	3576
ORB			
30541	32052	39404	24751
23299	37132	32453	21529
16438	19599	16178	11117

Figura 4-36 : Número de keypoints malos según cuadrante vídeo puente

Los datos que se acaban de mostrar difieren un poco con su gráfica de totales y la correspondiente con el número de keypoints buenos, que entre sí son iguales. Los extractores SIFT y ORB coinciden con el cuadrante de máximos, pero el de mínimos cambia, siendo la región marcada en la gráfica de totales 4-30 la esquina inferior izquierda y aquí se ha marcado la derecha. Para SIFT, se debería coger el 5º mínimo, mientras que para ORB debe ser el 3º. En el caso del extractor SURF, todo es diferente aunque también coindice con la casilla de mínimo de los otros extractores. Para que el máximo sea igual debería marcarse la región número 3, que en realidad es el segundo máximo del cuadrante, pero para el caso del mínimo su posición no es tan cercana, debiéndose marcar el 6º.

A continuación, se muestra la figura 4-37 que recoge los datos de keypoints buenos obtenidos en la cancha de baloncesto.

SIFT			
6616	7546	8835	7352
3105	4104	4597	5807
2052	2331	3092	4483
SURF			
7875	7584	9593	8017
3909	4183	5688	7019
1534	2951	4330	5782
ORB			
15760	17251	31658	18393
14913	17076	24843	25999
6657	15568	16599	17872

Figura 4-37 : Número de keypoints malos según cuadrante vídeo cancha baloncesto

En la imagen anterior se observa que los extractores coinciden, aunque no lo hacen con la figura 4-31 que recoge el número total de keypoints, que además es igual a la gráfica que recoge el número de keypoints correctos. La igualdad entre las imágenes se tiene en el cuadrante donde ocurre el máximo, pero el mínimo ocurre en la esquina inferior izquierda en vez de la derecha. En el caso de SIFT, se debería coger el 6º mínimo, para SURF el 7º y con ORB el 8º, que está más cerca del máximo que del mínimo.

Para concluir este apartado, se muestra la última grafica (4-38) en la que aparece la información recogida de keypoints incorrectos en el vídeo de pista de karts.

SIFT			
4721	4872	3976	2670
3735	5303	4896	2502
4477	5878	6708	3649
SURF			
5296	5608	4367	3386
6128	7612	7143	4153
7104	9071	10145	7903
ORB			
3930	3667	3190	1275
4595	8210	7396	2931
2970	7497	6798	3345

Figura 4-38 : Número de keypoints malos según cuadrante vídeo pista karts

Aunque los diferentes extractores no coinciden entre sí, cada uno de ellos marca las mismas casillas que se marcaron en la gráfica de totales (4-32). A pesar de coincidir con su gráfica total, hay algunos cambios respecto a la imagen donde se recoge el número de keypoints buenos (4-35). En su análisis se recoge las diferencias que aquí vuelven a ser las mismas.

4.3 Resultados del análisis de asociación de features

En este apartado se van a mostrar los resultados tras analizar los matches que ocurren durante la ejecución de los diferentes videos. Se recuerda que un match es la unión entre dos puntos pertenecientes a diferentes fotos, buscando que tengan características idénticas; es así como se diferencia a los matches buenos y malos, en función de si el enlace ha resultado correcto o no. La búsqueda de matches funciona de tal forma que la foto primera siempre va a buscar el análogo a todos los puntos clave. Por ello, las gráficas relacionadas con la búsqueda de máximos y mínimos ya han sido calculadas en el apartado 4.2, en el que se analizaban, entre otros, el máximo global de buenos, el mínimo global de puntos etc.

1. Número de matches

Se comienza este nuevo subapartado mostrando en la figura 4-39 el número total de matches que se han calculado para cada uno de los métodos de estudio y los diferentes vídeos. Se ha tenido en cuenta para los extractores SIFT y ORB que existen coordenadas repetidas, y por tanto han sido eliminadas del estudio. Las gráficas que se muestran a continuación son muy similares a las ya mostradas cuando se estudió el número total de keypoints (4-3); esto es debido a que los matches se realizan para todos los puntos existentes, excepto los de la última foto que no tienen pareja para compararse.

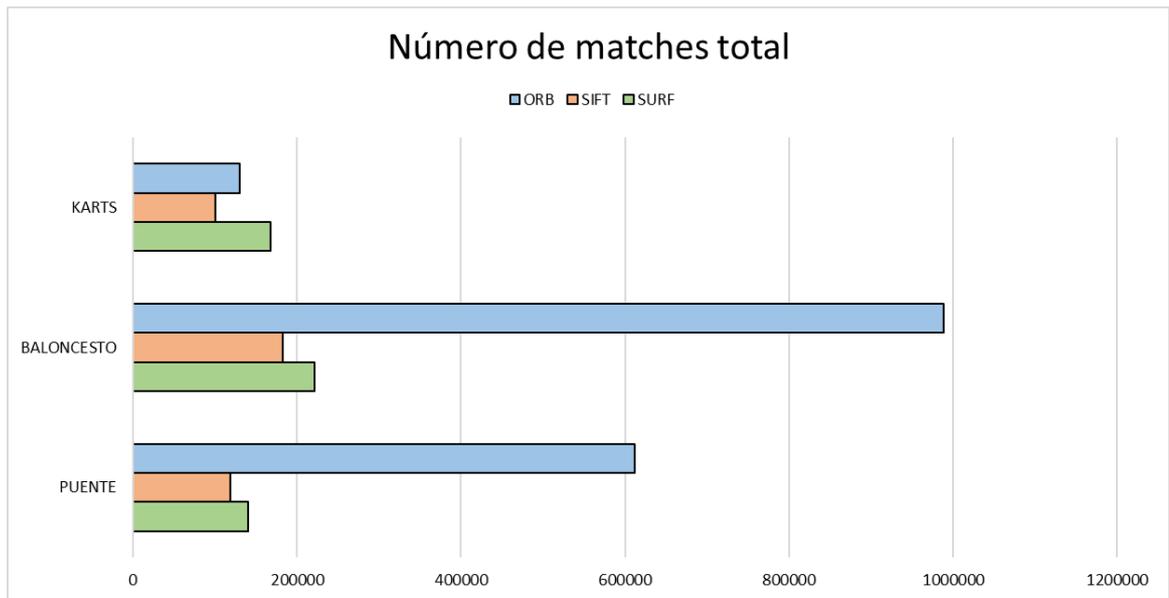


Figura 4-39 : Gráfica número de matches total

Como se puede observar, los resultados son prácticamente iguales a la imagen 4-3. Con la metodología ORB se obtienen una gran cantidad de matches, excepto en el vídeo karts. A pesar de lo resultados, lo importante es obtener una tecnología que sea capaz de calcular un gran número de matches buenos, estando pareja esta idea con obtener gran cantidad de matches.

2. Número de matches buenos

Se pasa ahora a estudiar la cantidad de enlaces correctos obtenidos en la gráfica 4-40.

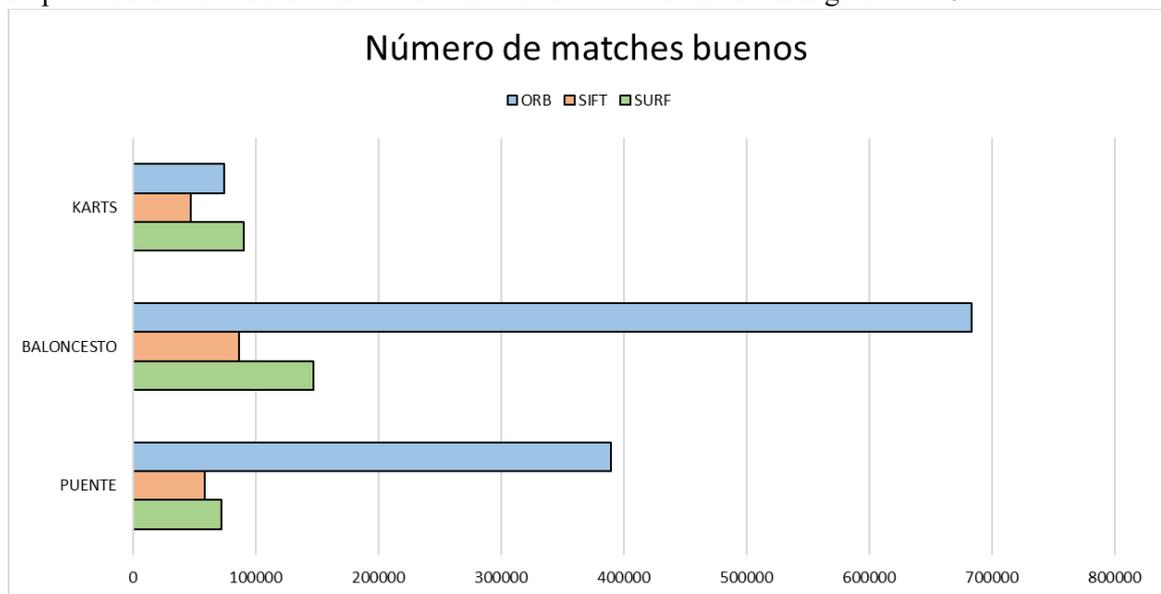


Figura 4-40 : Gráfica número de matches buenos

Los resultados son muy similares a la gráfica del número total de matches, siendo de nuevo ORB el que ocupa el primer lugar seguido de SURF y SIFT. Lo que llama la atención es que la situación se repite excepto en karts, donde ORB desciende a la segunda posición. Para entender mejor si el número de matches buenos es adecuado en función del total, se pasa a mostrar en la tabla 4-6 el porcentaje de enlaces adecuados en comparación con el total.

	Puente (%)	Baloncesto (%)	Karts (%)
SIFT	49.50	47.31	46.98
SURF	51.91	66.31	53.71
ORB	63.64	69.19	57.16

Tabla 4-6 : Porcentaje de matches buenos

Con la tabla anterior se puede concretar las ideas planteadas anteriormente. Como se puede ver, con ORB se logra un alto porcentaje de matches correctos, teniendo incluso un alto número de keypoints. El extractor con el peor porcentaje resulta ser SIFT, que apenas alcanza ni la mitad de los puntos.

3. Número de matches malos

Se pasa ahora a exponer en la gráfica 4-41 los resultados tras analizar los matches incorrectos.

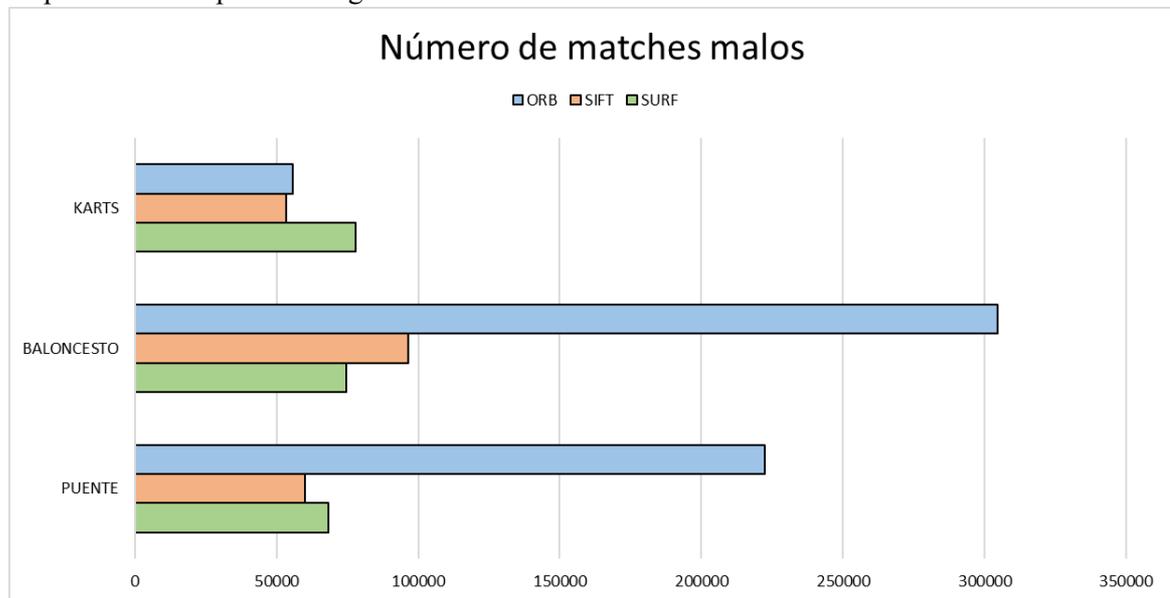


Figura 4-41 : Gráfica número de matches malos

El extractor ORB resulta ocupar la primera posición en cuanto a cantidad de matches incorrectos. Como se ha comentado anteriormente, se aprecia mejor estos resultados con una tabla (4-7) de porcentajes.

	Puente (%)	Baloncesto (%)	Karts (%)
SIFT	50.50	52.69	53.02
SURF	48.69	33.69	46.29
ORB	36.36	30.81	42.84

Tabla 4-7 : Porcentaje de matches malos

El objetivo es trabajar con un extractor que tenga un bajo porcentaje de enlaces incorrectos. Como se aprecia en la tabla anterior, con ORB se consigue este hito, sobrepasando apenas que un tercio de los datos sean incorrectos.

4. Número de matches según cuadrante

En este subapartado se ha estudiado la procedencia del cada match. Para ello, se ha calculado a que cuadrante pertenece cada uno de los extremos del enlace. En el caso de que los números sean diferentes, se tomará como match entre cuadrantes diferentes. En primer lugar, se pasa a clasificar los matches totales para cada uno de los vídeos, siendo el primero el grabado en el puente, que corresponde a la figura 4-42.

SIFT				
8328	9087	9790	8191	Match entre cuadrantes diferentes
5818	7344	6144	6868	
3317	4431	5398	6452	
SURF				
9467	8957	10646	8731	Match entre cuadrantes diferentes
7249	8128	8410	9198	
3438	6863	8639	8932	
ORB				
32064	43958	60833	36823	Match entre cuadrantes diferentes
40282	53985	50861	59259	
19330	43297	40353	35205	

Figura 4-42 : Número de matches total según cuadrante vídeo puente

En esta ocasión se aprecia que los extractores coinciden en los máximos y mínimos. Cuando se ha usado ORB se observa que los números son los más altos, incluso se obtiene mayor cantidad de match entre cuadrantes diferentes; el extractor con los datos más bajos es SIFT. Además, también coinciden en las posiciones con la gráfica que recoge el número de keypoints según cuadrante (figura 4-30).

En la ilustración 4-43 se muestran los datos correspondientes al vídeo de la cancha del baloncesto.

SIFT				
16434	16772	19132	17231	Match entre cuadrantes diferentes
5669	8515	8520	7292	
6132	6591	6380	3831	
SURF				
28341	26542	28559	23191	Match entre cuadrantes diferentes
9307	12164	11595	11677	
6946	7953	7510	5782	
ORB				
67116	98352	128875	66923	Match entre cuadrantes diferentes
58607	119874	105085	52572	
35594	49622	38303	23034	

Figura 4-43 : Número de matches total según cuadrante vídeo baloncesto

Al igual que antes, las regiones marcadas son la misma para las tres tecnologías, incluso coinciden otra vez con la gráfica de keypoints (4-31). El primer lugar en comparación con los otros extractores está ocupado por ORB, incluyendo el mayor número de matches entre cuadrantes diferentes.

Para finalizar con el número total de matches, se muestra en la imagen siguiente (4-44) los resultados del vídeo tomado en la pista de karts.

SIFT				
8181	9511	6750	4542	Match entre cuadrantes diferentes
7361	9550	8340	5309	
8835	17292	17823	12627	
SURF				
13703	14225	17540	13834	Match entre cuadrantes diferentes
9693	11088	12463	12969	
4341	8414	10858	11612	
ORB				
7769	7761	6264	1804	Match entre cuadrantes diferentes
7557	15854	14835	3829	
3167	10231	8881	2737	

Figura 4-44 : Número de matches total según cuadrante vídeo karts

Cuando se estudió el número de keypoints totales en la pista de karts (4-32), pasó lo mismo que ahora; los extractores no coinciden como ha ocurrido con los dos vídeos anteriores. A pesar de ello, los tres coinciden que la región donde ocurren menos matches se encuentra en una esquina, aunque difieren en la localización. En cuanto al lugar donde ocurren más enlaces, son totalmente diferentes. Además, no se pudo encontrar una relación cercana, ya que por ejemplo para que coincida ORB con SIFT debería estar marcado el cuarto máximo.

Comienza una nueva tanda de tres gráficas que recogen los matches buenos, mostrándose primero en la siguiente figura (4-45) los datos recogidos en el puente.

SIFT				
4941	5854	6282	4989	Match entre cuadrantes diferentes
4456	5703	4650	4178	
2203	3304	3926	3947	
SURF				
5277	5539	6741	5232	Match entre cuadrantes diferentes
5268	6122	5938	5469	
2674	5155	6085	5468	
ORB				
20587	33583	43758	26183	Match entre cuadrantes diferentes
29523	42574	39107	41970	
15319	32442	30891	23277	

Figura 4-45 : Número de matches buenos según cuadrante vídeo puente

Se observa que las tres tecnologías coinciden, además que también lo hacen con la figura 4-42 que recoge los matches totales. Esta situación también ocurrió cuando se estudió el número de keypoints buenos según cuadrantes en el vídeo del puente.

A continuación, en la imagen 4-46 se muestra el resultado del vídeo de la cancha de baloncesto.

SIFT				
9730	11735	12555	9345	Match entre cuadrantes diferentes 60691
4040	6195	6450	5032	
3628	4492	3983	2336	
SURF				
20110	21054	22589	17459	Match entre cuadrantes diferentes 41853
7281	10234	9851	9049	
4470	6099	5760	3922	
ORB				
47641	80682	105457	50686	Match entre cuadrantes diferentes 144283
45381	98865	87065	39806	
25798	39072	29288	16262	

Figura 4-46 : Número de matches buenos según cuadrante vídeo cancha baloncesto

Los tres extractores coinciden, además también lo hacen con la gráfica que recoge los totales de matches (4-43). Además, como se acaba de comentar, también ocurre como en el número de keypoints: las imágenes pertenecientes tanto al total como al número de puntos correctos coinciden en las posiciones de cuadrantes máximos y mínimos.

Para terminar el estudio de los cuadrantes correctos, se muestra en la imagen 4-47 los resultados tras analizar el vídeo de la pista de karting.

SIFT				
4624	4570	3592	1880	Match entre cuadrantes diferentes 37926
2555	3209	2501	1583	
3109	5987	6558	3458	
SURF				
5898	7385	5088	3254	Match entre cuadrantes diferentes 52196
4667	6192	5290	3780	
5845	13784	13534	9609	
ORB				
5395	6027	4973	1281	Match entre cuadrantes diferentes 39572
5179	12168	11056	2560	
1901	6866	5809	1165	

Figura 4-47 : Número de matches buenos según cuadrante vídeo pista karts

En la figura anterior se observa que ninguno de los extractores coincide entre sí. Si se observa su gráfica predecesora en la que se recogen el número total de matches (4-44), sólo se obtienen 2 coincidencias:

en el máximo de SIFT y en el de ORB. Fijando la vista en ORB, para que los datos coincidan no es muy difícil, ya que se marca el cuadrante 4 que en este caso es el segundo mínimo de la gráfica. Con SIFT, ocurre lo mismo que se acaba de comentar. El extractor que obtiene unos resultados más dispares es SURF, con el que cuesta más encontrar una relación con la figura 4-44.

Se pasa a exponer el último tramo de gráficas relacionadas con cuadrantes, en concreto se van a mostrar las relacionadas con la contabilización de matches incorrectos. La primera que se muestra es la figura 4-48, que recoge la información del vídeo del puente.

SIFT				
3387	3233	3508	3202	Match entre cuadrantes diferentes
1362	1641	1494	2690	
1114	1127	1472	2505	
SURF				
4190	3418	3905	3499	Match entre cuadrantes diferentes
1981	2006	2472	3729	
764	1708	2554	3464	
ORB				
11477	10375	17075	10640	Match entre cuadrantes diferentes
10759	11411	11754	17289	
4011	10855	9462	11928	

Figura 4-48 : Número de matches malos según cuadrante vídeo puente

En este caso, el único extractor que coincide tanto con su gráfica de total (4-42) y gráfica de matches buenos (4-45) es SIFT, aunque ORB y SURF están muy próximos a sus resultados anteriores; ORB debería marcar como máxima la región número 3 que es el segundo máximo, y SIFT marca el cuadrante 1 que es también el segundo máximo.

A continuación (4-49), se recogen los matches malos según cuadrantes en el vídeo de la cancha de baloncesto.

SIFT				
6704	5037	6577	7886	Match entre cuadrantes diferentes
1629	2320	2070	2260	
2504	2099	2397	1495	
SURF				
8231	5488	5970	5732	Match entre cuadrantes diferentes
2026	1930	1744	2628	
2476	1854	1750	1860	
ORB				
19475	17670	23418	16237	Match entre cuadrantes diferentes
13226	21009	18020	12766	
9796	10550	9015	6772	

Figura 4-49 : Número matches malos según cuadrante vídeo cancha baloncesto

Observando las figuras 4-43 y 4-46 (gráficas de totales y de matches buenos respectivamente), se observan que son iguales entre sí, y la figura anterior no lo es. Se comienza con SIFT: los mínimos coinciden, aunque para que esto ocurra con los máximos, se debe coger el tercer máximo. Continuando el análisis con SURF, el mínimo marcado es el segundo, cosa que también ocurre con el máximo. Por último, ORB coincide con el mínimo y el máximo, además que es el extractor que obtiene mayor número de matches entre cuadrantes diferentes.

Para finalizar el análisis de matches incorrectos, y de todo este apartado que analiza la asociación de features, se muestra en la figura 4-50 los datos pertenecientes al vídeo de la pista de karts.

SIFT				
1592	1575	1195	961	Match entre cuadrantes diferentes
1378	2052	1956	801	
1734	1911	2741	1235	
SURF				
2283	2126	1662	1288	Match entre cuadrantes diferentes
2694	3358	3050	1529	
2990	3508	4289	3018	
ORB				
2374	1734	1291	523	Match entre cuadrantes diferentes
2378	3686	3779	1269	
1266	3365	3072	1572	

Figura 4-50 : Número matches malos según cuadrante vídeo pista karts

Observando la gráfica de totales (4-44) y la de matches buenos (4-47), se concluyó que no son muy coincidentes, aspecto que vuelve a repetirse; cuando se analizaron los cuadrantes de los keypoints ocurrió lo mismo. El extractor SIFT coincide en sus tres máximos, y su mínimo coincide con su predecesora de matches correctos, por lo que de nuevo es el segundo mínimo. Con SURF, se tiene el mismo mínimo que ocurre en 4-47, que para que sea igual a su gráfica de totales se debería marcar el cuadrante 9, que es el séptimo mínimo. Para que coincida con el máximo de 4-44, se ha de marcar la región número 3, que se encuentra muy lejana para ser un máximo. Por último, analizando ORB, coincide con su mínimo global, aunque de nuevo su máximo es el segundo en comparación con su gráfica de totales (4-44).

4.4 Resultados del análisis por condiciones de contraste

En el análisis del contraste se ha recurrido a estudiar los diferentes cuadrantes y su comportamiento antes condiciones lumínicas. Para ello, se han realizado dos estudios: calcular el contraste para cada uno de los cuadrantes y otro que se encarga de saber la cantidad de keypoints según el contraste calculado. Estos análisis sirven para conocer el comportamiento de la aparición de puntos clave en función de las condiciones de brillo y contraste del entorno, es decir, se desea conocer la influencia de la luz en la aparición o no de puntos de información.

1. Contraste de cada cuadrante

Se exponen a continuación en estudio, para cada uno de los vídeos, del contraste obtenido. Para recordar su cálculo se recomienda leer la página 11. En dicho apartado, se calculó el contraste para cada vídeo completo, que a diferencia de las gráficas que se van a mostrar analizan el contraste según el cuadrante;

Por ello, las gráficas pueden resultar algo diferentes. En primer lugar, en la figura 4-51 se analiza el contraste del vídeo grabado en el puente.

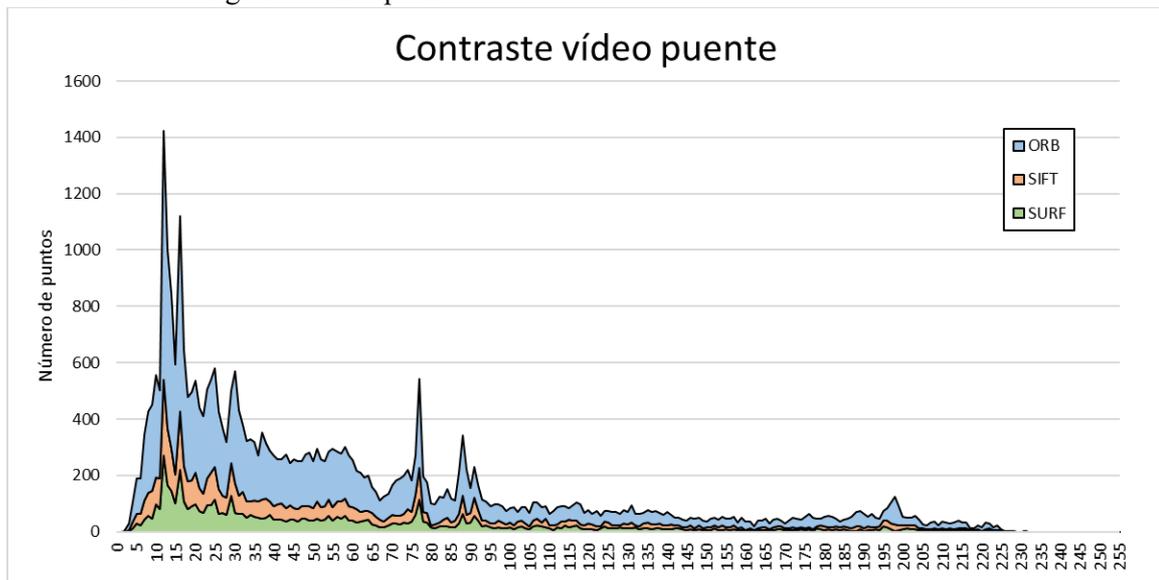


Figura 4-51 : Contraste según cuadrantes vídeo puente

En primer lugar, se observa que ORB obtiene mucha mayor área, lo que se traduce en mayor número de datos. En segundo lugar, las gráficas son prácticamente iguales, aunque no lo son con la figura 3-3 que analiza el contraste total del vídeo; se debe a que en este apartado se estudia el contraste según el cuadrante, no total. En la figura anterior, se observa que se obtienen valores de contraste muy bajos, obteniéndose un pico en valores en torno a 0 (valores muy oscuros). Se concluye que analizando los cuadrantes se aporta nueva información en comparación con el vídeo completo, ya que en la figura 3-3 se concluyó que el vídeo estaba bien contrastado, al contrario que aquí que se observa que está poco contrastado, por lo que existen muchos cuadrantes con tonalidades oscuras que hacen que en esta gráfica se concluya que el vídeo está subexpuesto.

A continuación, la figura 4-52 muestra el resultado del vídeo de la cancha de baloncesto, tras analizar el valor del contraste para cada uno de sus 12 cuadrantes.

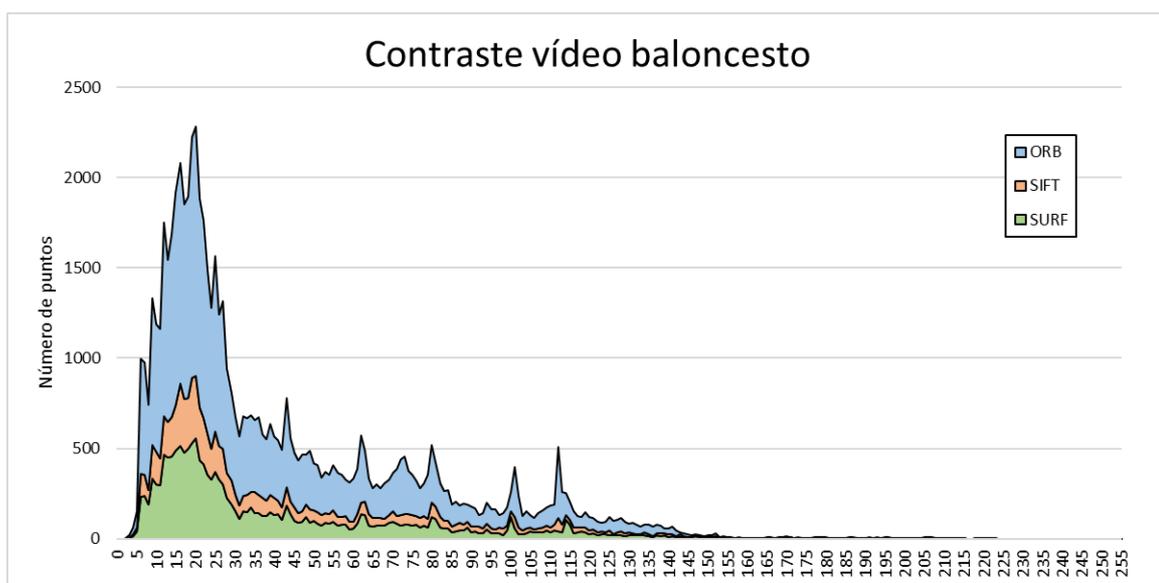


Figura 4-52 : Contraste según cuadrantes vídeo baloncesto

Al igual que la figura perteneciente al vídeo del puente, se observa que con ORB se obtienen muchos más datos en comparación con los demás. De nuevo, se produce un pico en la parte izquierda de la gráfica, lo que indica que el vídeo está compuesto por un gran número de imágenes oscuras. Si analizamos la imagen 3-6, ambas gráficas en esta ocasión son muy similares. En aquella ocasión se llegó a la conclusión que es un vídeo subexpuesto, como ocurre también analizando los cuadrantes. Por ello, se concluye que no existe una gran pérdida de información al analizar los cuadrantes en vez del vídeo completo, aunque teniendo en cuenta los cuadrantes se obtiene un fuerte pico en la parte izquierda, lo que quiere decir que aportan mucha oscuridad al conjunto total del vídeo.

Por último en este subapartado, se analiza el vídeo de karts, mostrándose sus resultados en la figura 4-53.

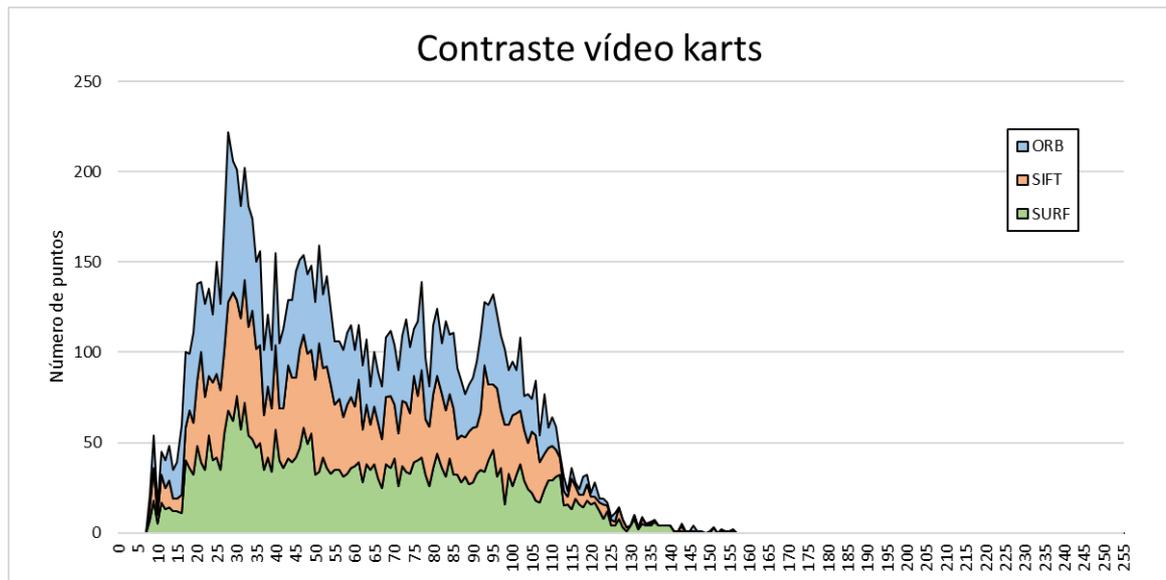


Figura 4-53 : Contraste según cuadrantes vídeo karts

Recurriendo de nuevo al análisis del contraste del vídeo completo (3-9), se puede encontrar una gran similitud entre ambas figuras. Se produce un claro allanamiento de la parte derecha, concluyéndose que el vídeo se encuentra subexpuesto, es decir, con gran parte de las fotos que lo componen con tonos oscuros. Además, se llega a la conclusión que, a pesar de no analizar el vídeo completo, no se pierde tanta información, lográndose una gran similitud entre las gráficas.

2. Keypoints según contraste

En este subapartado se estudia de nuevo el contraste según el cuadrante, aunque esta vez se acumulan para cada resultado el número de keypoints que se encuentran en dicho cuadrante. Lo que se quiere es conocer como es capaz de encontrar cada extractor puntos clave, en función de si la imagen es oscura o clara.

En primer lugar, se ha analizado el vídeo grabado en el puente, mostrándose el resultado en la figura 4-54.

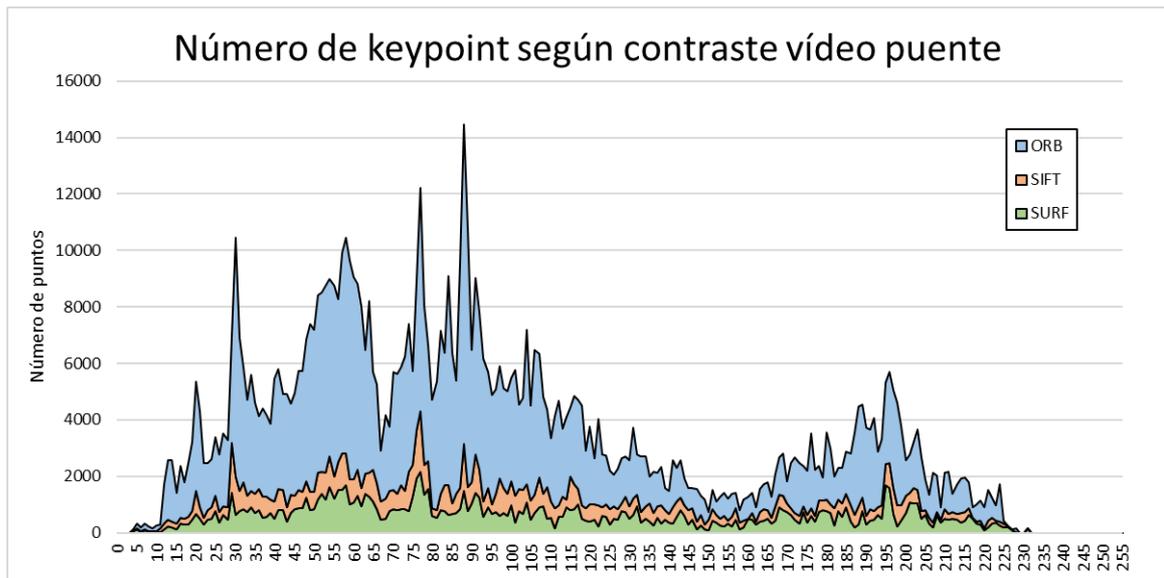


Figura 4-54 : Gráfica número de keypoints según contraste vídeo puente

Esta gráfica tiene sus datos muy distribuidos a lo largo de ella, aunque se aprecia un cúmulo claro en la parte izquierda de la figura, en la que se marcan los puntos oscuros. Además, si se compara con el estudio del contraste del vídeo completo (3-3), se parecen mucho las gráficas, cosa que no ocurría al analizar el contraste según el cuadrante (4-51). Volviendo de nuevo a la gráfica que nos acontece, la amplia distribución de los puntos a lo largo de la figura indica que en este caso al calcular keypoints no influye la luz, aunque es más fácil encontrarlos cuando la imagen tiene colores semioscuros.

La figura siguiente (4-55), muestra el análisis de keypoints según el contraste para el vídeo grabado en la cancha de baloncesto.

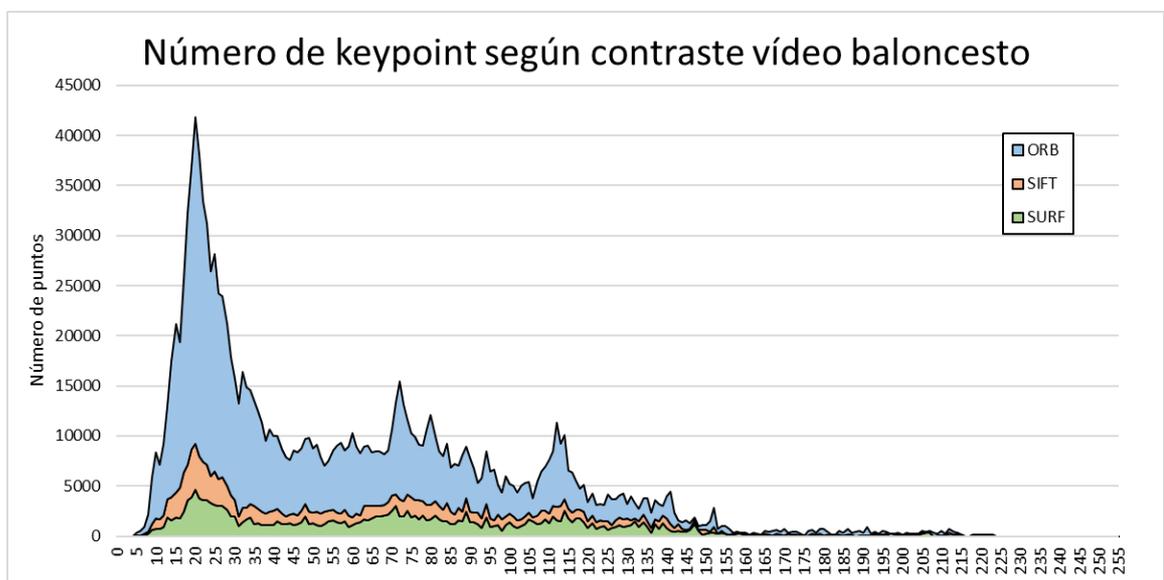


Figura 4-55 : Gráfica número de keypoints según contraste vídeo baloncesto

En esta ocasión se muestra un claro pico en la parte izquierda de la gráfica, con valores muy cercanos a cero. Además, una vez ocurrido ese pico, la gráfica comienza a aplanarse hasta tal punto que en la mitad derecha de la gráfica, perteneciente a valores claros, apenas hay datos. Con ello se concluye que en este vídeo es más sencillo encontrar puntos clave cuando las condiciones lumínicas son bajas, y cuando se tiene mucha luz apenas se consigue encontrar keypoints.

Para finalizar, se muestra el estudio realizado en la pista de karts, mostrándose dicho resultado en la figura 4-56.

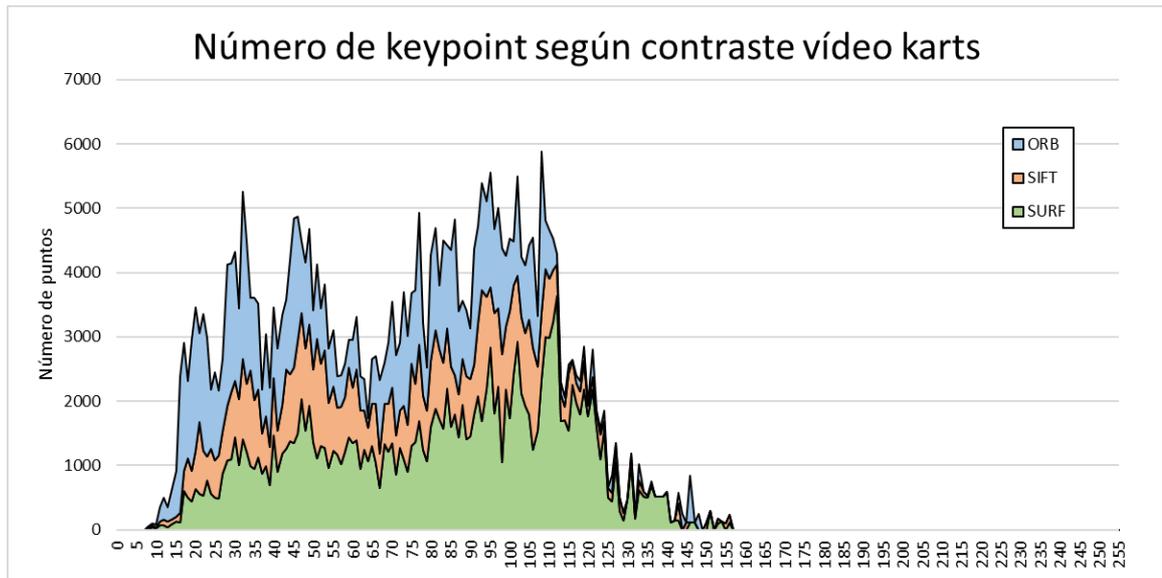


Figura 4-56 : Gráfica número de keypoints según contraste vídeo karts

Como se puede observar, las tres áreas son muy parecidas, por lo que los extractores obtienen más o menos la misma cantidad de datos, aunque de nuevo ORB sigue siendo el que resalta más. Si se divide la gráfica en dos, se aprecia que el conjunto de datos se acumula claramente en la mitad oscura (0-128), obteniéndose una distribución casi uniforme, sin picos muy remarcables. Por ello, se llega a la conclusión de que en este vídeo se es capaz de calcular más keypoints en zonas oscuras y de medio color, y que cuando las imágenes comienzan a aclararse la posibilidad de obtener puntos clave es prácticamente nula.

4.5 Resultados de costes computacionales

Para medir la eficacia computacional de cada una de las tecnologías que se están estudiando, se han llevado a cabo tres análisis: contabilizar el número de fotos, medir el tiempo de cálculo de los keypoints y el de los matches. Con estas características lo que se busca es conocer al extractor más rápido, aunque siempre se necesita que sea rápido, eficaz y veraz.

1. Número de fotos

En este primer estudio se han contado el número de fotos que es capaz de calcular cada tecnología. Para ello, se muestra en la figura 4-57 dichos resultados.

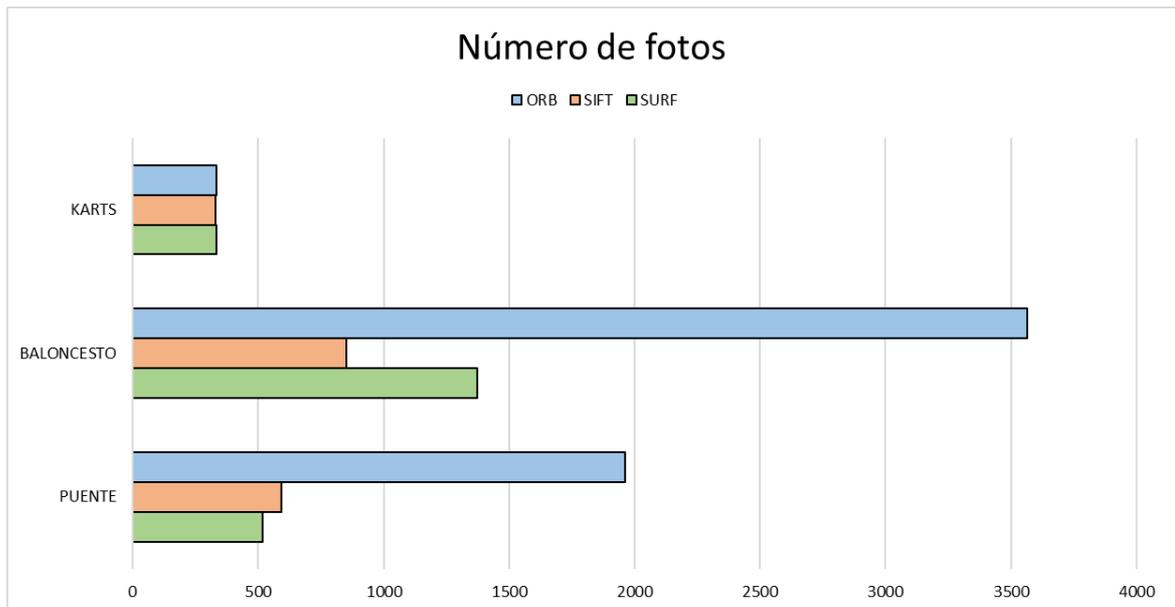


Figura 4-57 : Gráfica número de fotos

Como se puede observar ORB es con diferencia el extractor que obtiene más fotos en el tiempo de ejecución de los vídeos, por lo que obtendrá más información durante el funcionamiento del programa. En el video karts se puede apreciar que todos los extractores están muy parejos.

En la tabla a continuación (4-8) se muestra el porcentaje de fotos que cada uno de los extractores es capaz de calcular, en comparación con el número total de fotos que componen cada uno de los videos. Dicha información aparece en la descripción de cada uno de los bags, dentro del capítulo 3.2.

	Puente (%)	Baloncesto (%)	Karts (%)
SIFT	19.38	14.25	98.20
SURF	16.92	22.96	99.40
ORB	64.38	59.61	99.40

Tabla 4-8 : Porcentaje del número de fotos de cada vídeo respecto al total

De esta tabla se puede observar varias cosas. La primera de ella es que del vídeo karts los tres extractores son capaces de analizar casi todas las fotos. La segunda es que SIFT y SURF obtienen porcentajes parecidos y muy bajos, por lo que apenas obtienen información de los vídeos. Por último, se concluye que ORB es un buen extractor en el sentido de optimizar los recursos que puede aprovechar de cada uno de los vídeos, siendo capaz de analizar aproximadamente el 60% de la totalidad del vídeo.

2. Tiempo de cálculo de keypoints

En este subapartado se van a analizar los tiempos de cálculo de keypoints: máximo, mínimo y medio. Se parte inicialmente con la figura 4-58, que muestra el resultado de tiempos máximos.

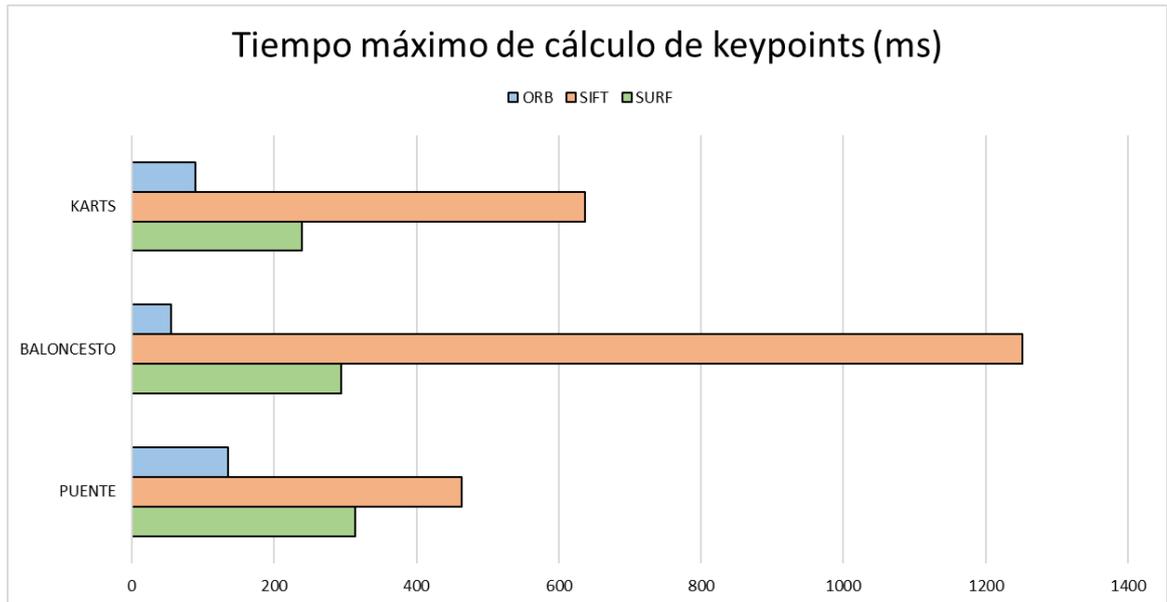


Figura 4-58 : Gráfica tiempo calculo keypoints máximo (ms)

Se persigue conseguir que los tiempos de ejecución sean bajos, que se logra principalmente con la tecnología ORB. El extractor SIFT alcanza unos máximos muy altos en comparación con el resto, observándose incluso que en la cancha de baloncesto llega a tardar más de 1 segundo en calcular los puntos clave.

A continuación, se muestra en la figura 4-59 los tiempos de cálculo mínimo.

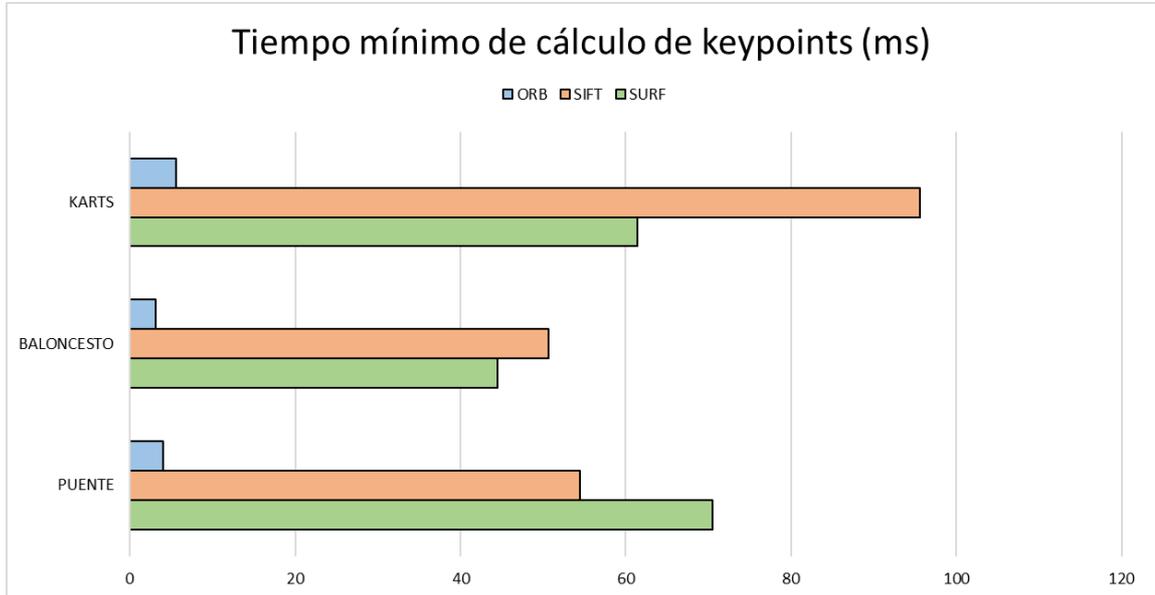


Figura 4-59 : Gráfica tiempo cálculo keypoints mínimo (ms)

Al igual que antes, se persigue obtener un extractor con tiempos de cálculo mínimo bajos, que de nuevo se repite para el extractor ORB. El resto de las tecnologías se ven muy parejas, excepto en el vídeo de la pista de karting, donde se observa que SIFT obtiene una gran diferencia en comparación con el resto.

Debido a que con los tiempos máximos y mínimos no se obtiene información totalmente distribuida de la realidad, se pasa a exponer a continuación diferentes gráficas que recogen los tiempos de cálculo de keypoints para cada extractor y vídeo. En dichas gráficas se puede obtener tanto el máximo, el mínimo, la mediana (que indica el punto donde se acumula el 50% de los datos) y la media.

En primer lugar, en la figura 4-60 se muestran los datos de los tiempos recogidos en el vídeo del puente, para el extractor SIFT.

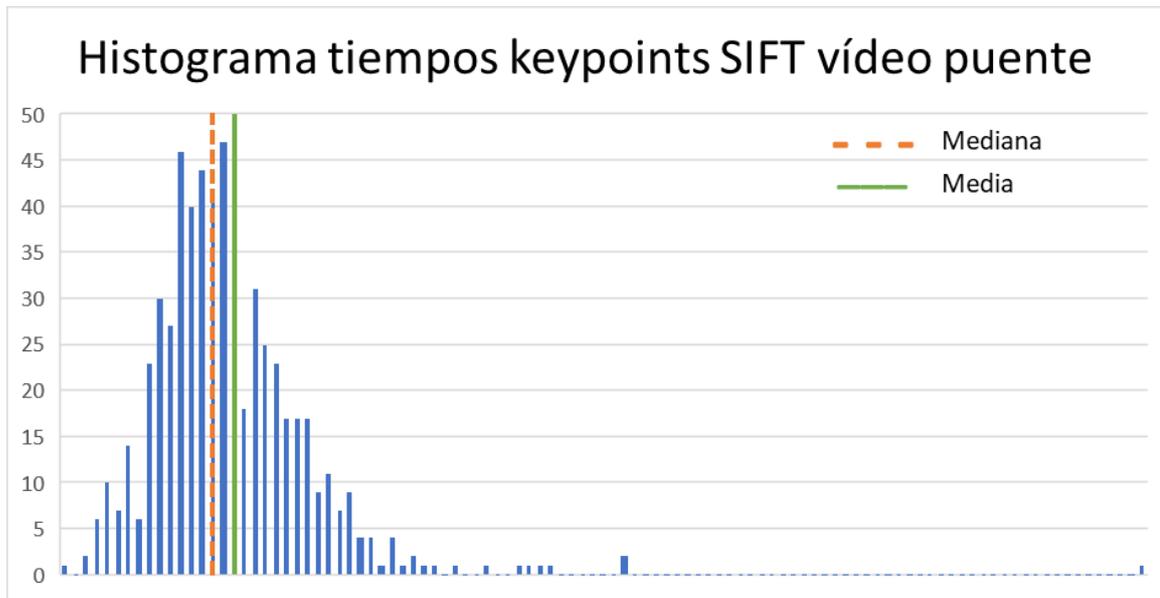


Figura 4-60 : Gráfica tiempos cálculo keypoints SIFT video puente

Se puede observar que con los tiempos anteriores se obtiene un pico en la parte más cercana al mínimo, y en el resto de la gráfica apenas se aprecian muchos datos. Esto es debido a que existen algunas ocasiones que se tarda un poco más en calcular los keypoints, aunque se puede ver que apenas ocurre con mucha frecuencia. La media y la mediana ocurren en la parte izquierda de la gráfica, lo que indica que los tiempos de cálculo de puntos clave están muy cercanos al mínimo.

A continuación, en la imagen 4-61 se recogen los tiempos obtenidos en la cancha de baloncesto para SIFT.

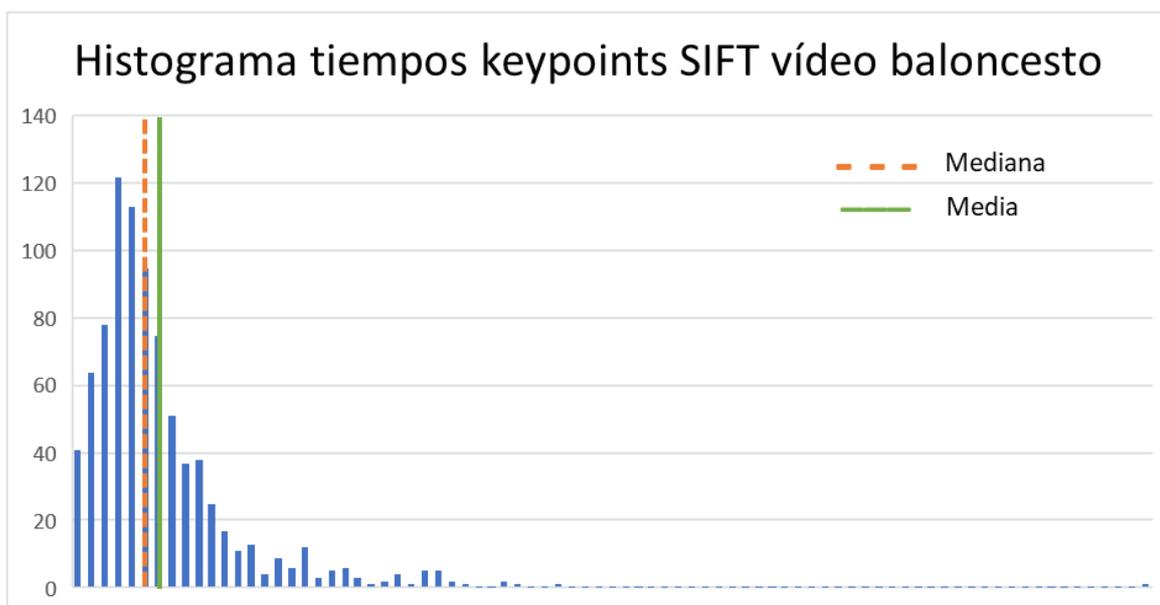


Figura 4-61 : Gráfica tiempos cálculo keypoints SIFT video baloncesto

Al igual que antes, se observa como los datos se acumulan en la parte izquierda, dejando la parte derecha prácticamente plana. La media y mediana vuelven a casi coincidir, además de que ambas se encuentran

cerca del mínimo. Debido a la disposición de los datos, se puede concluir que los tiempos de cálculo son muy pequeños, en torno al mínimo.

Por último, para el extractor SIFT se expone la gráfica 4-62 que recoge los tiempos de cálculo de keypoints en el vídeo de la pista de karts.

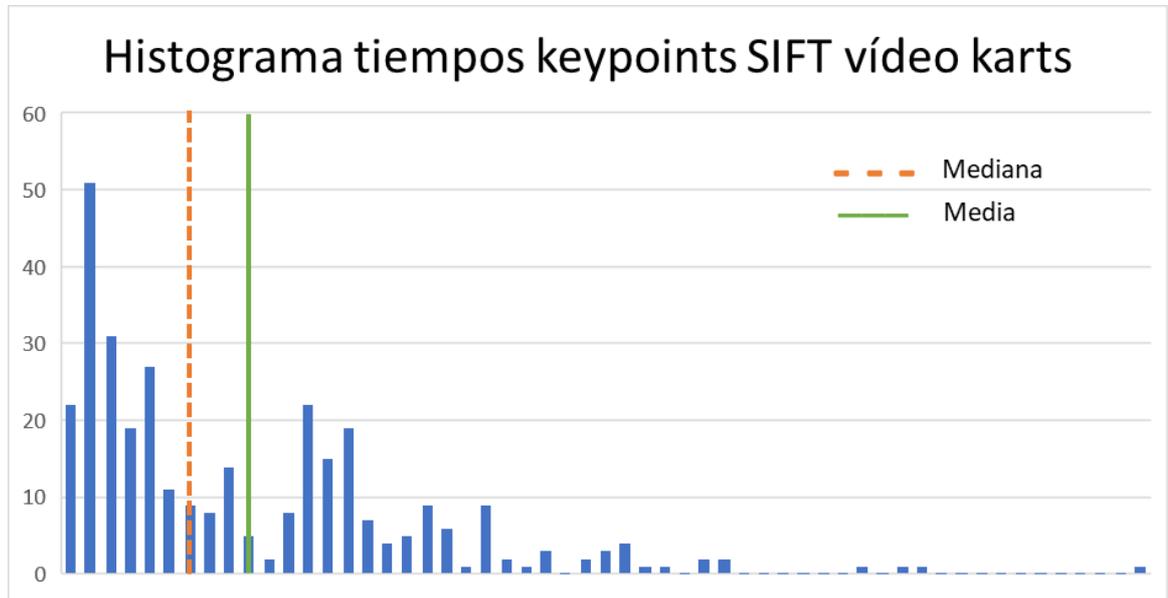


Figura 4-62 : Gráfica tiempos cálculo keypoints SIFT video karts

En este histograma no se nota una clara campana sesgada como ha ocurrido en las dos graficas anteriores. Además, la diferencia entre la mediana y la media es más notoria, encontrándose la media a la derecha. Esto indica que la distribución de los tiempos tiende más a alejarse del mínimo, y que para calcular puntos clave se tarda un poco más.

Se exponen a continuación 3 nuevas gráficas para el extractor SUFT, siendo la primera de ellas la 4-63, que recoge los datos en el bag puente.

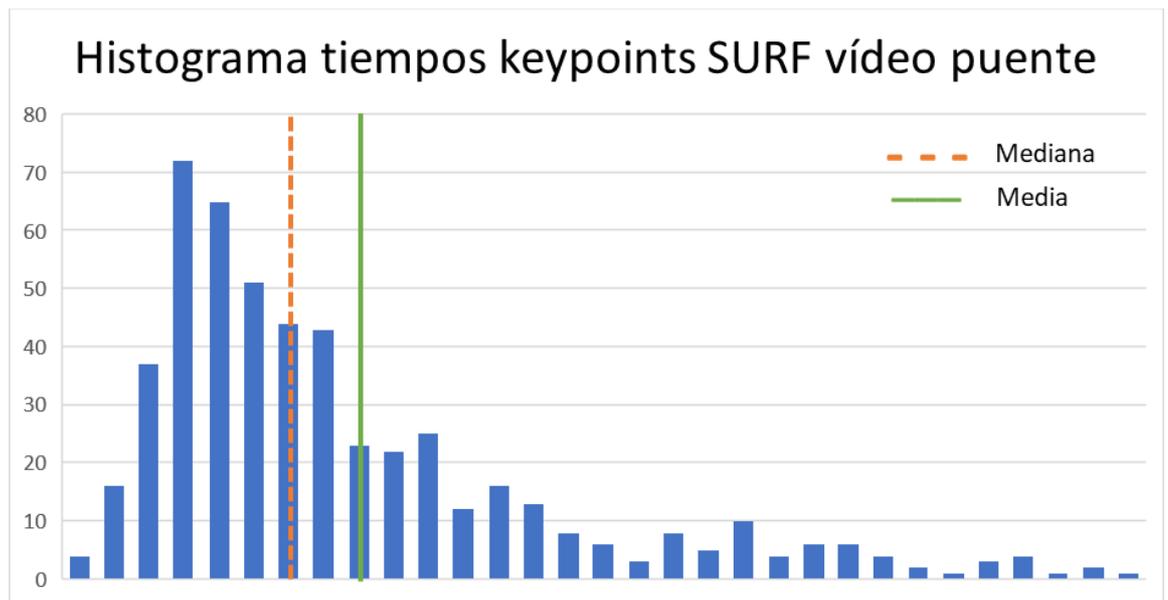


Figura 4-63 : Gráfica tiempos cálculo keypoints SURF video puente

En esta ocasión, se puede ver que tiene forma de campana sesgada a la izquierda, y que incluso los datos

se encuentran más repartidos a lo largo de la gráfica. A pesar de ello, la acumulación de datos se produce en gran parte en la mitad izquierda de la gráfica, y que la media de tiempos comienza a acercarse a los valores más altos. Por todo lo anterior, se puede decir que la media de cálculo de keypoints es baja, aunque comienza a alejarse de los tiempos más bajos.

La figura siguiente (4-64), muestra los datos temporales del extractor SURF en la cancha de baloncesto.

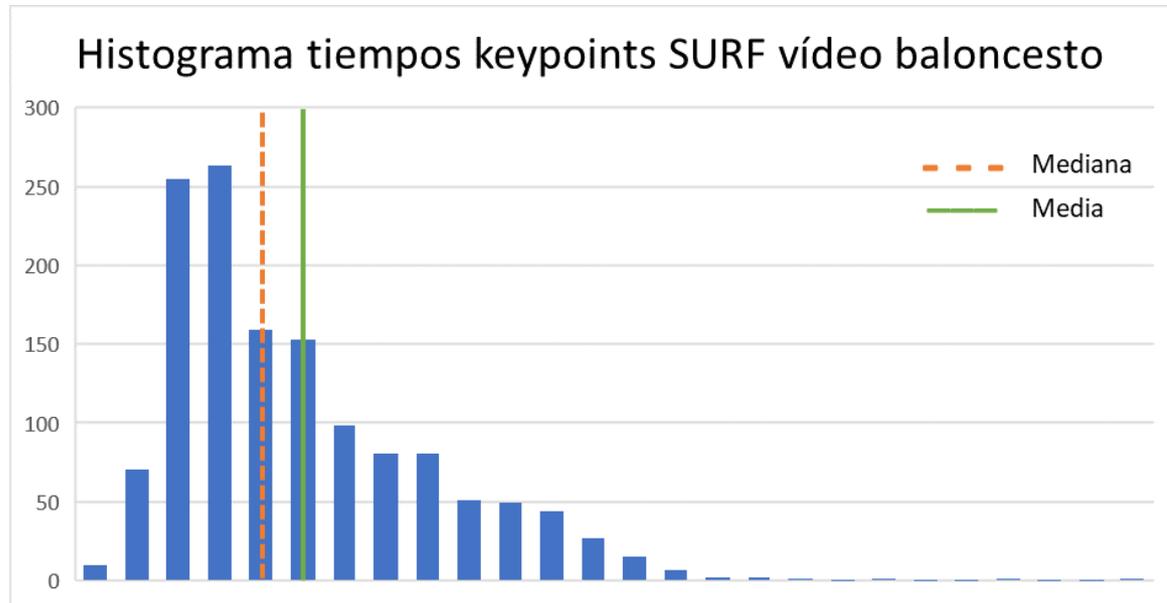


Figura 4-64 : Gráfica tiempos cálculo keypoints SURF video baloncesto

En la figura anterior se ve cómo se acumula toda la información al comienzo de la gráfica, para después decaer lentamente. Se puede apreciar que apenas existen datos en la mitad derecha, provocado porque existen ocasiones en los que se calculan tiempos más grandes en comparación con el resto. La media y la mediana están muy cercanas, colocadas en la parte de valores de tiempos pequeños, aunque por la posición de la media respecto a la mediana se puede decir que los tiempos comienzan a crecer.

Por último, se muestra en la gráfica 4-65 los tiempos obtenidos por SURF en la pista de karts.

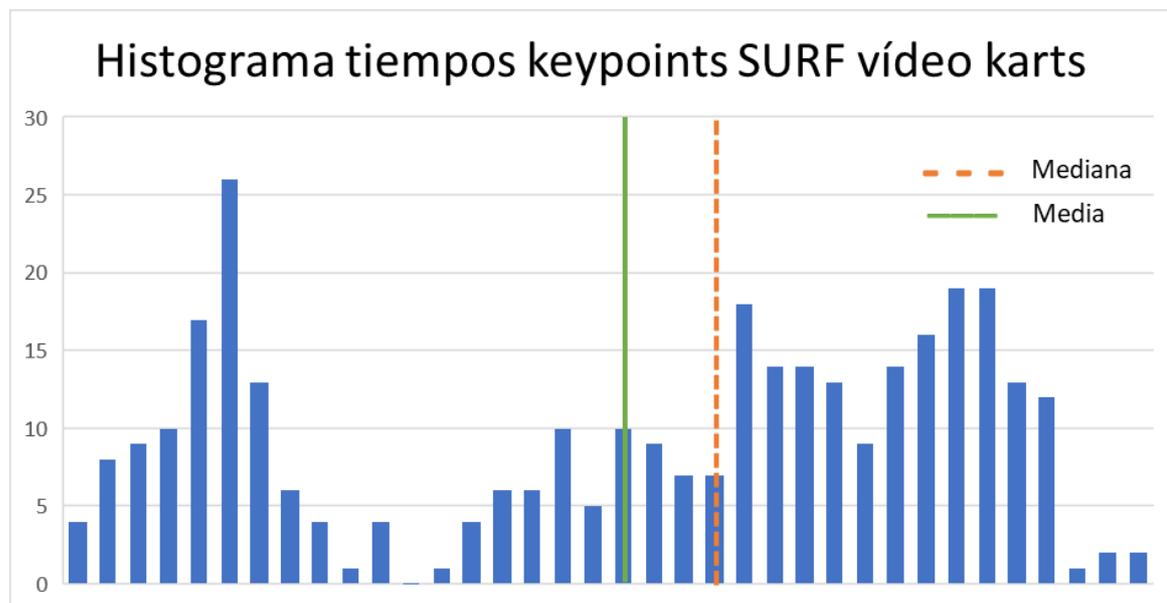


Figura 4-65 : Gráfica tiempos cálculo keypoints SURF video karts

Se observa que esta gráfica no tiene ninguna forma característica, distribuyéndose la información por todo el espacio de la figura. A pesar de encontrarse un fuerte pico casi al comienzo, debido a la distribución de los datos la mediana se encuentra en la mitad derecha de los datos. Además, se observa que la media es mucho más pequeña que la mediana, así que a pesar de que los datos comiencen a aglomerarse en los valores más altos, la media nos indica que dichos resultados están distribuidos y existen valores pequeños de tiempos.

Se presenta a continuación las últimas tres gráficas de este subapartado, pertenecientes a la tecnología ORB. Se comienza con la gráfica 4-66 que recoge los datos temporales en el vídeo del puente.



Figura 4-66 : Gráfica tiempos cálculo keypoints ORB video puente

El histograma anterior muestra cómo se obtiene un fuerte pico al comienzo de la gráfica, y que después se comienza a tender a cero. Esto es debido a que existen algunas ocasiones en los que los valores temporales obtenidos son muy altos en comparación con los demás, pero debido a la gran cantidad de datos dichos valores son insignificantes. La media de los datos es superior a la mediana, aunque al ser ambas pequeñas y similares, se concluye que ORB obtiene tiempos de cálculo de keypoints pequeños.

La siguiente figura (4-67) recoge la información temporal de keypoints en la pista de baloncesto.

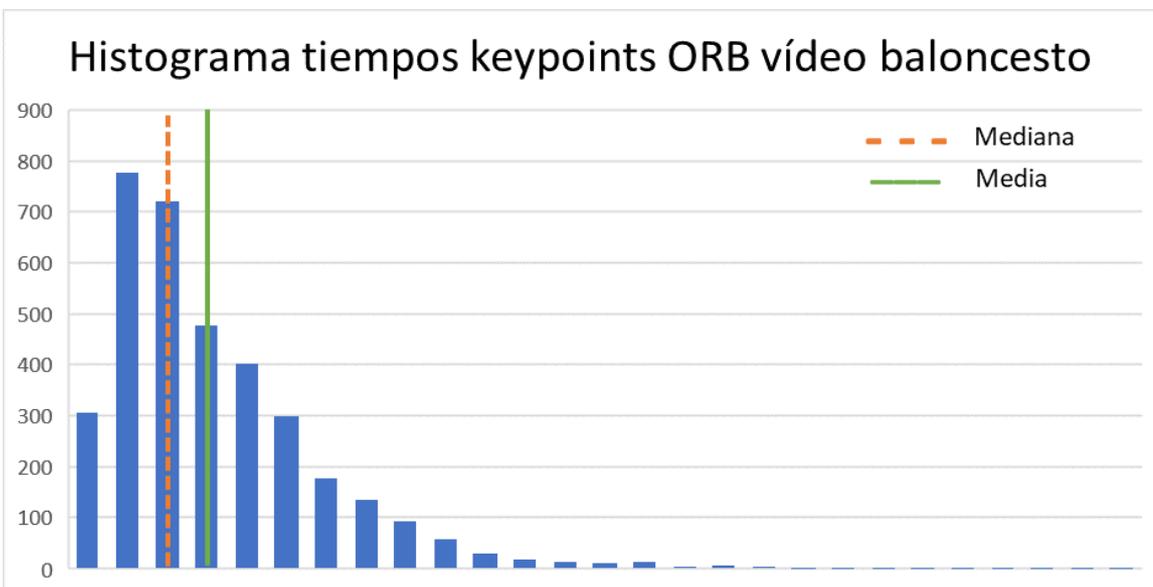


Figura 4-67 : Gráfica tiempos cálculo keypoints ORB video baloncesto

Se observa de nuevo una clara tendencia a cero en los valores más altos, aunque esta vez el pico no está tan marcado como antes. De nuevo, mediana y media son pequeñas y están muy cercanas entre sí, lo que indica que los tiempos para calcular keypoints son pequeños, aunque existan algunas fotográficas en las que se tarde un poco más en calcular dicha información.

Para concluir con ORB, se muestra su última gráfica (4-68), que recoge los tiempos en la pista de karting.

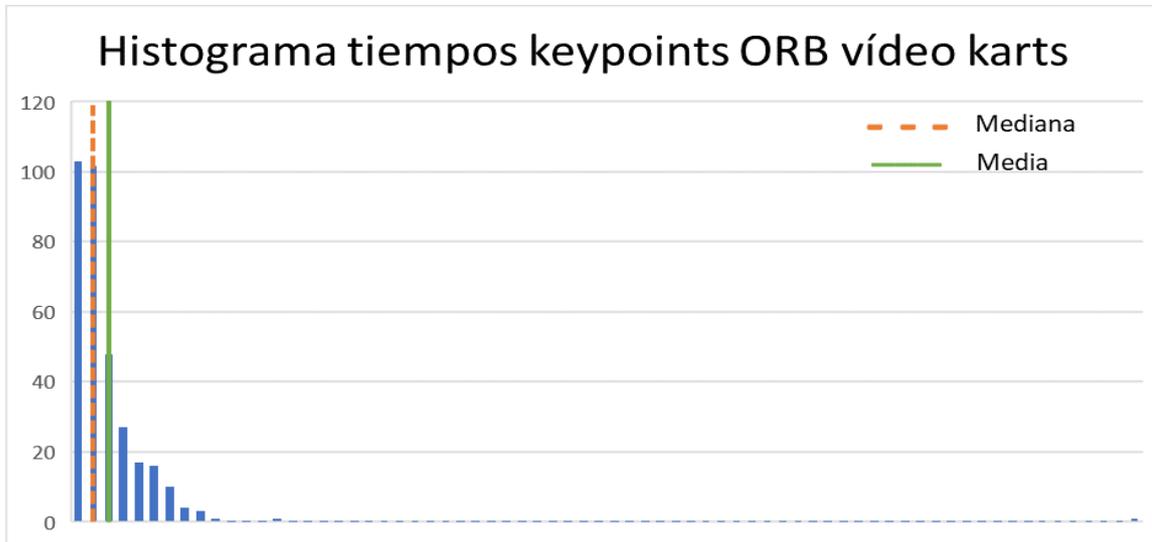


Figura 4-68 : Gráfica tiempos cálculo keypoints ORB video karts

En todas las gráficas obtenidas para ORB se observa siempre la misma forma: un pico a la derecha y una larga tendencia en torno al cero, con la mediana y la media con valores similares y pequeños. Esta situación se repite en la gráfica actual, por lo que de nuevo se concluye que para calcular keypoints ORB emplea poco tiempo.

3. Tiempo cálculo de matches

En este apartado se pasa a mostrar los tiempos de cálculo de los enlace entre fotos: máximos, mínimos y medios. En primer lugar, se muestra la gráfica 4-69 que recoge los tiempos máximos.

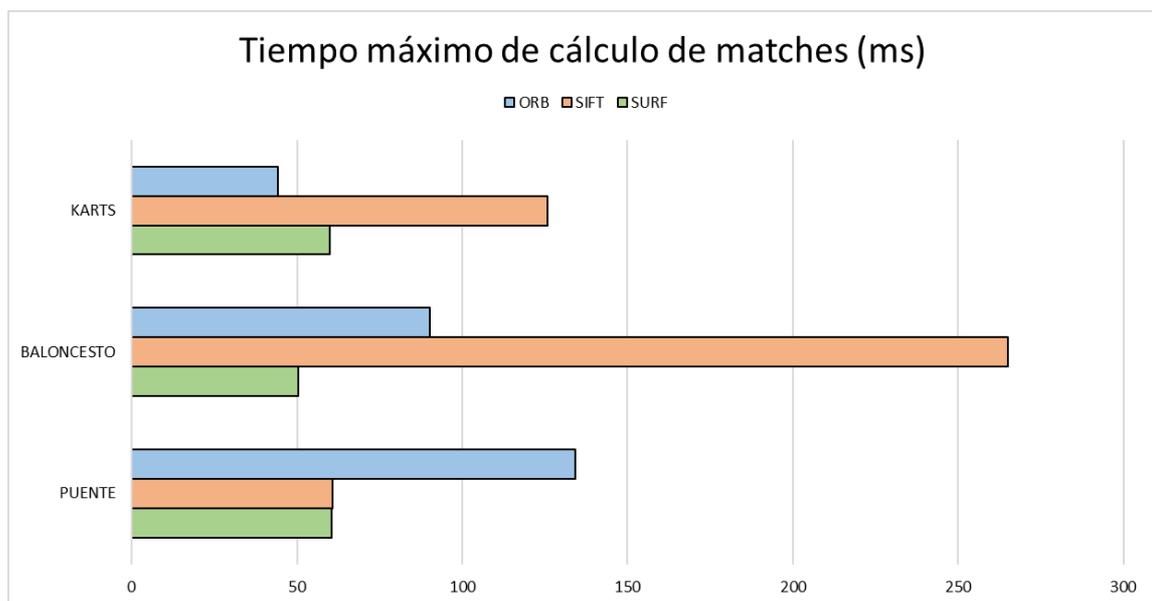


Figura 4-69 : Gráfica tiempo cálculo matches máximo (ms)

Se aprecia que gran parte de los datos se acumula en la parte izquierda de la gráfica, confirmándose dicho hecho con la aparición de la mediana al comienzo de la figura. La media de los datos temporales es mayor que la mediana, por lo que se llega a que con SIFT en este bag se obtiene una media de tiempos de matching que comienza a ser elevada, aunque sigue estando cerca del mínimo.

El siguiente histograma (4-72) recoge la información del bag de la pista de baloncesto.

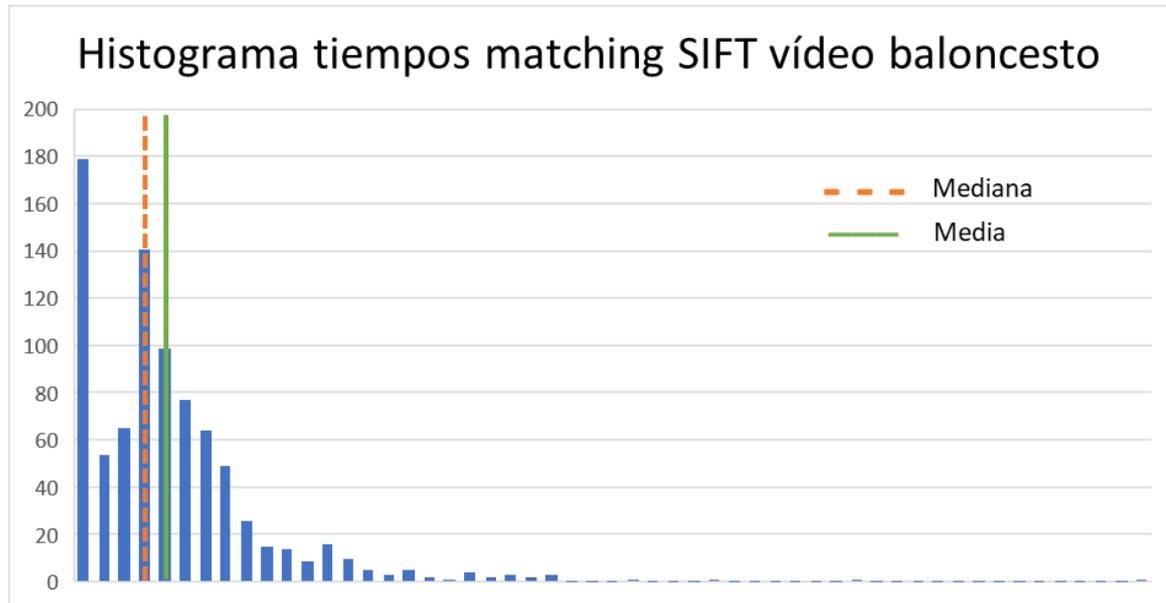


Figura 4-72 : Gráfica tiempos cálculo matching SIFT video baloncesto

En la gráfica anterior se puede apreciar una fuerte acumulación de los datos en los tiempos más bajos, obteniéndose una media y mediana muy similares. Además, existe una amplia zona de valores de tiempo elevados, aunque su cantidad es irrelevante. Por todo lo anterior se concluye que la media de tiempos tiende prácticamente a valores mínimos, por lo que para realizar matches se emplea poco tiempo en este vídeo y con este extractor.

Para acabar con esta tecnología, se muestran los resultados obtenidos en el vídeo de karting en la imagen 4-73.

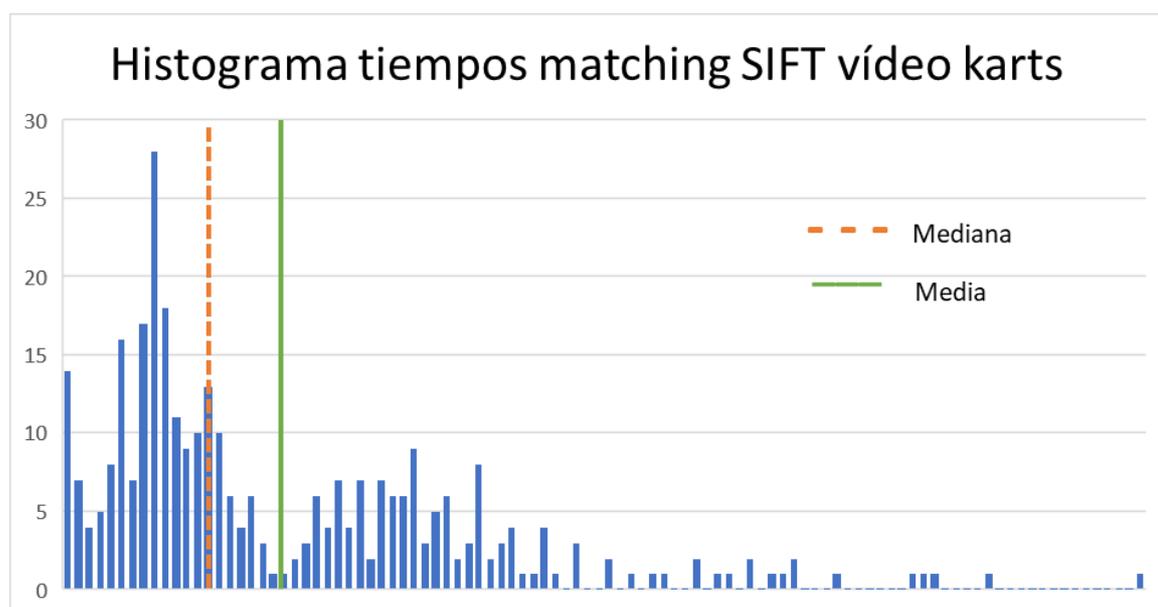


Figura 4-73 : Gráfica tiempos cálculo matching SIFT video karts

La grafica anterior no muestra ninguna forma reconocida, aunque gracias a la ayuda de los marcadores se puede entender mucho mejor el histograma. Debido al tamaño de las barras se puede concretar que existe un gran número de datos, y además la gran mayoría de ellos se encuentra en la zona de valores más bajos, información obtenida con la mediana que marca el punto donde se acumula el 50% de los datos. Además, la línea marcada en verde muestra donde se encuentra la media de todos los tiempos calculados, llegándose a concluir que a pesar de que se tiene un gran número de valores altos, la media es cercana al mínimo de tiempo, por lo que para este vídeo se tarda poco en realizar enlaces aunque comienza a crecer su valor.

Las tres nuevas gráficas que aparecen a continuación recogen la información obtenida con SURF para los diferentes video, siendo el primero el grabado en el puente (figura 4-74).

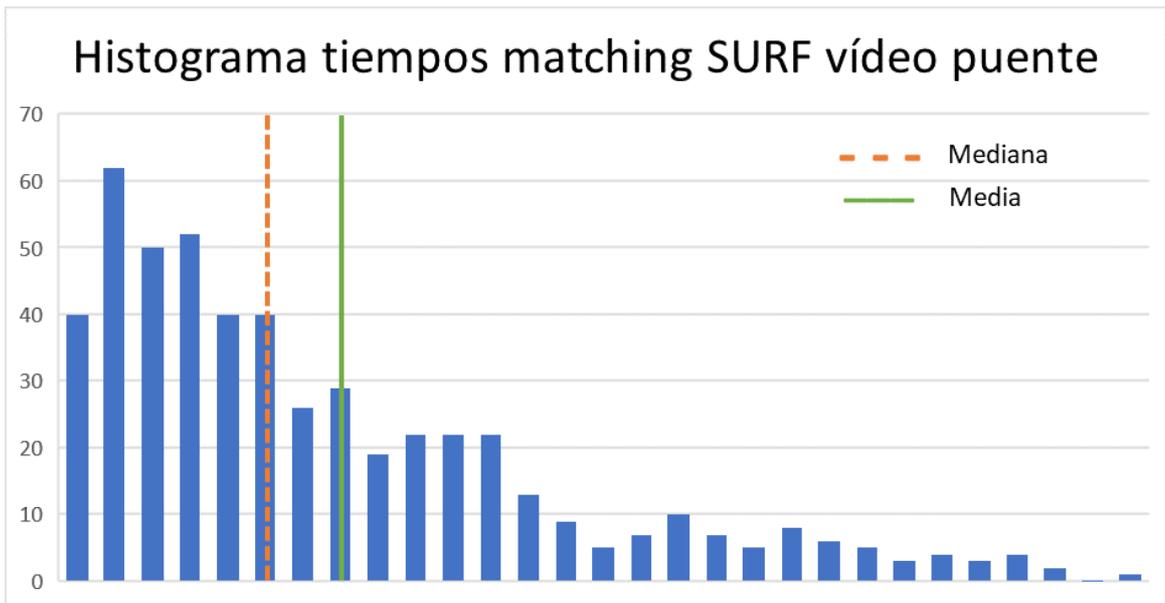


Figura 4-74 : Gráfica tiempos cálculo matching SURF video puente

Se observa que, debido a la posición de la mediana más del 50% de los datos ocupan la mitad derecha de la gráfica. Además, en esta zona se encuentra señalada la media, por lo que los tiempos de matching de SURF para este vídeo son favorables, aunque comienzan a alejarse del mínimo.

La siguiente figura (4-75), muestra los tiempos obtenidos en la cancha de baloncesto.

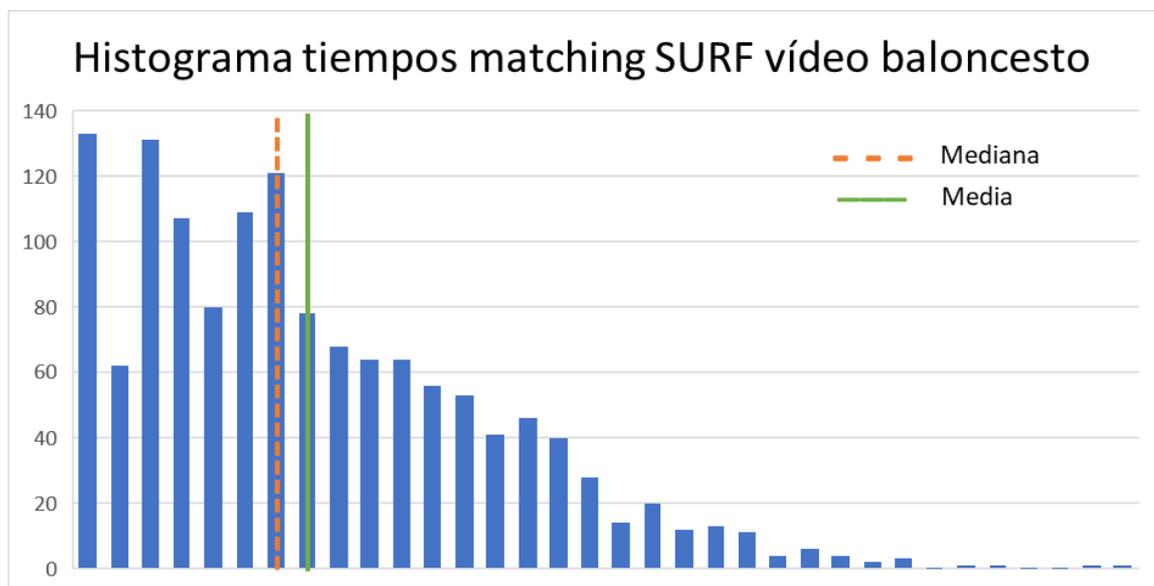


Figura 4-75 : Gráfica tiempos cálculo matching SURF video baloncesto

Se observa que, debido a que la mediana está situada a la izquierda y cercana al mínimo, la otra mitad de datos tiene una gran extensión hasta el final de la gráfica. La media en este caso está situada a la derecha de la mediana, aunque muy próximas. Por ello, la asociación de features tiene unos tiempos que comienzan a alejarse de lo deseado, que en este caso es tener tiempos bajos.

Para concluir con SURF, se expone su última gráfica (4-76), que recoge la información en el bag de la pista de karting.

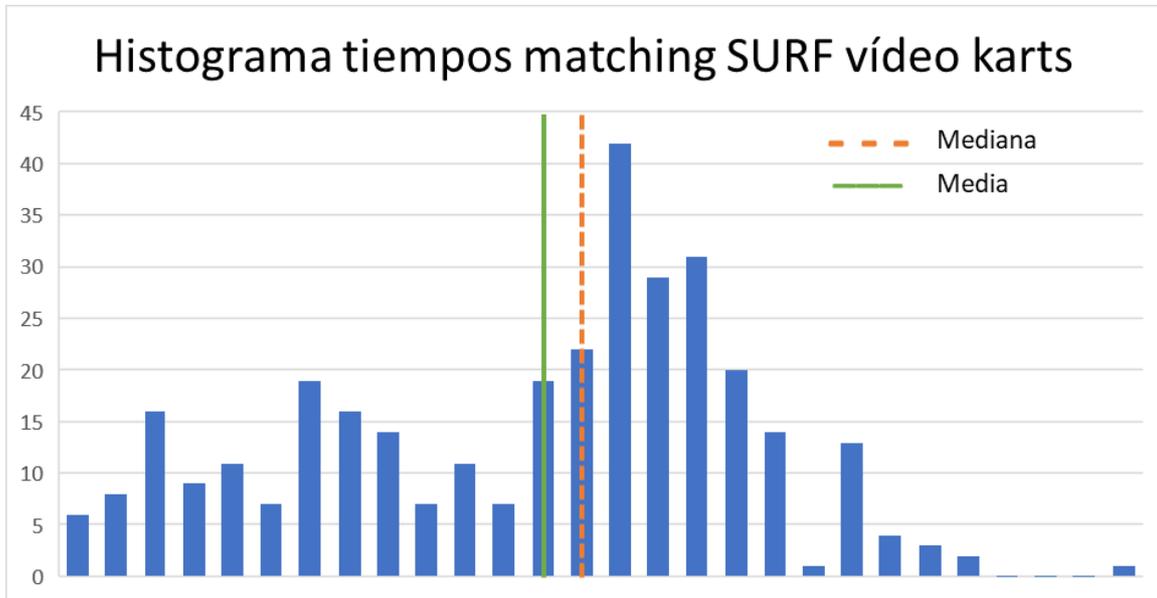


Figura 4-76 : Gráfica tiempos cálculo matching SURF video karts

Como se observa, la mediana y la media están justo en la mitad del gráfico; esto indica que los tiempos están distribuidos. Para concretar mejor, la media se encuentra en la zona más cercana a los mínimos, aunque muy centrada, por lo que para SURF en este vídeo no se tiene un tiempo medio de cálculo bajo.

A continuación, se pasa a exponer las últimas 3 gráficas que recogen los datos de ORB, siendo la primera de ellas la figura 4-77 que recopila la información en el vídeo del puente.

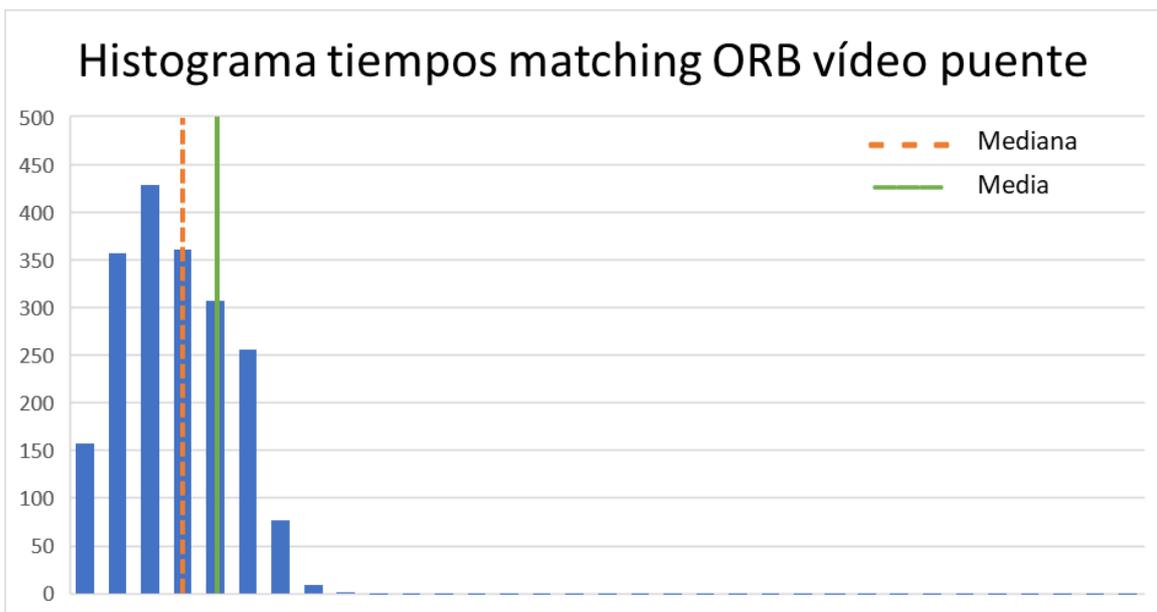


Figura 4-77 : Gráfica tiempos cálculo matching ORB video puente

Se puede ver que la extensión de datos desde la mediana hacia la derecha es muy amplia y con poco relevancia. Se añade además que la mediana y la media están muy cerca y desplazadas a la izquierda. Por todo ello se puede decir que para calcular matching en este vídeo se emplea poco tiempo, que es lo que se desea.

La siguiente imagen (4-78) corresponde a los datos tomados en la cancha de baloncesto.

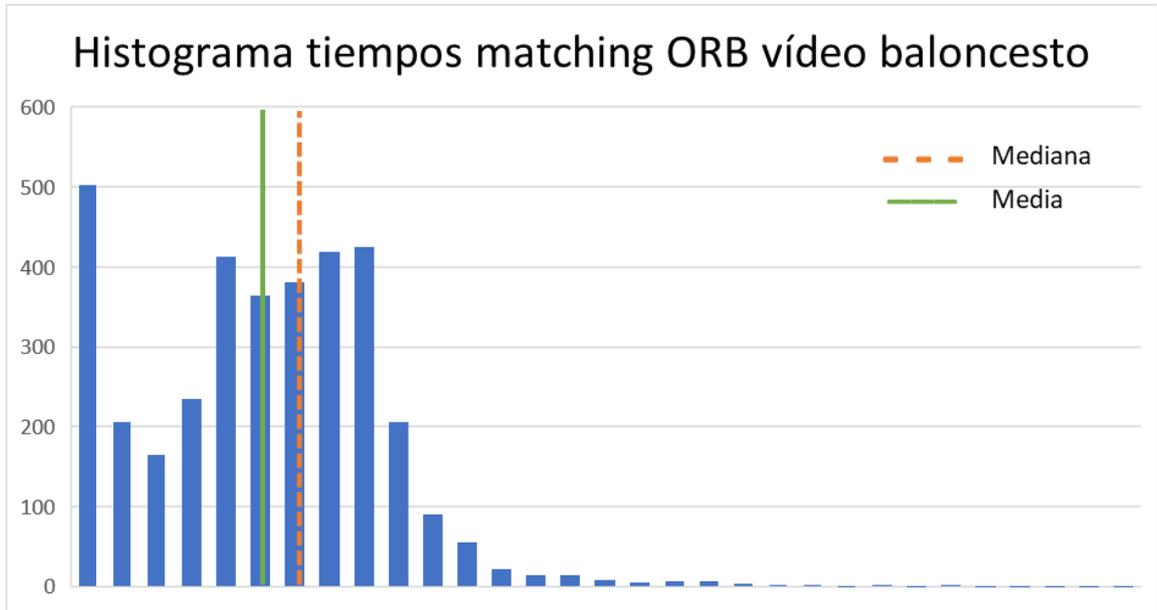


Figura 4-78 : Gráfica tiempos cálculo matching ORB video baloncesto

Se puede ver que mediana y media vuelven a estar muy juntas, y además en la parte más cercana al tiempo mínimo; incluso la media es menor que la mediana. Esto indica que de nuevo se necesitan pocos milisegundos para poder encontrar características similares entre dos fotos.

Por último, para finalizar con ORB y con el apartado en el que estamos, se muestra la última gráfica (4-79) que recoge la información de la pista de karting.

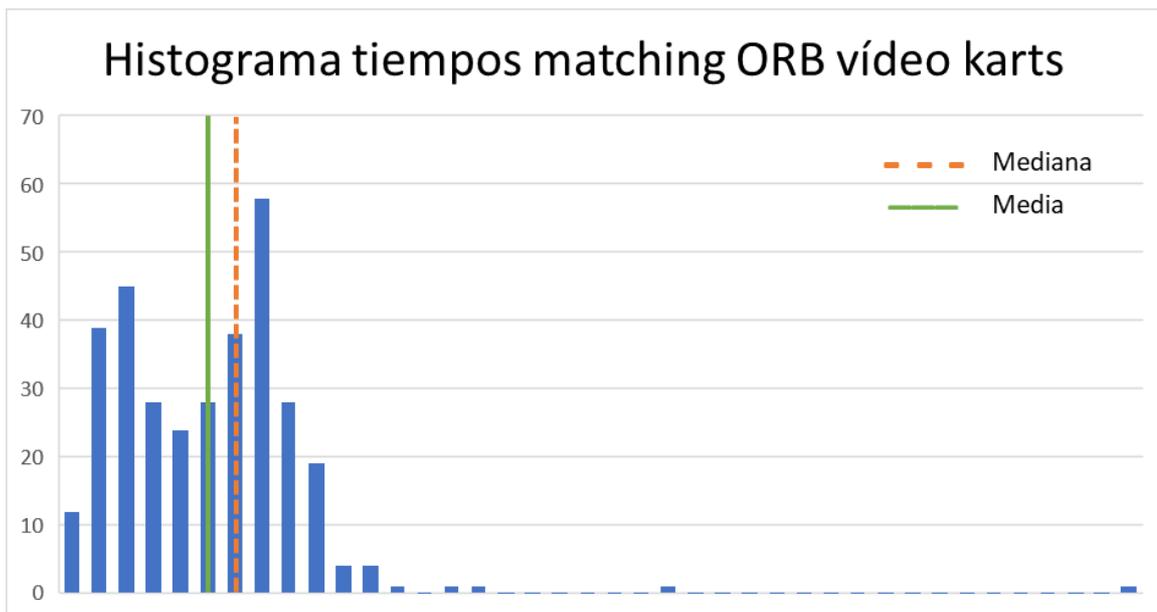


Figura 4-79 : Gráfica tiempos cálculo matching ORB video karts

Debido a la posición de la mediana, se observa que existe una extensión de datos muy amplia a la

derecha en la que se acumula el 50% de ellos, es decir, la mitad de los tiempos calculados cercanos al máximo hacen que la gráfica se amplíe el rango. Para conocer si los tiempos que se obtienen son favorables o no, se observa que la media de ellos se encuentra en la parte izquierda, muy cerca del mínimo, que es lo que interesa obtener.

4.6 Comparación de extractores

Una vez realizados todos los análisis, se pone ahora a disposición del lector una pequeña tabla resumen (4-9) de todo lo acontecido en este capítulo. Para cada una de las tecnologías estudiadas (SIFT, SURF y ORB) se señalarán aquellas características que han resultado relevantes e interesantes. En la tabla se disponen los nombres de los extractores de forma que a la izquierda se encuentra el más antiguo.

SIFT	SURF	ORB
<ul style="list-style-type: none"> - Más repetición de puntos - Menos keypoints - Menor % de keypoints buenos - Menor % de keypoints malos - Menor máx. de keypoints - Menor mín. de keypoints buenos - Mayor mín. de keypoints malos - Distancia media central, un poco separadas - Menor máx. y mín. de keypoints totales, buenos y malos según cuadrantes - Menos matches - Menor % de matches buenos - Mayor % de matches malos - Menor máx. y min de matches totales, buenos y malos según cuadrantes - Menos matches entre cuadrantes diferentes - Menos fotos analizadas - Mayor tiempo máx. de cálculo de keypoints - Mayor tiempo mín. de cálculo de keypoints - Mayor tiempo máx. de cálculo de matching - Mayor tiempo mín. de cálculo de matching 	<ul style="list-style-type: none"> - Repetición de puntos nula - Mayor % de keypoints malos - Mayor máx. de keypoints - Mayor máx. de keypoints buenos - Mayor máx. de keypoints malos - Menor mín. de keypoints - Menor mín. de keypoints malos - Distancia media central - Mayor tiempo medio de cálculo de keypoints - Menor tiempo máx. de cálculo de matching - Mayor tiempo medio de cálculo de matching 	<ul style="list-style-type: none"> - Más keypoints (2 de los 3 vídeos) - Mayor % de keypoints buenos - Menor máx. de keypoints malos - Mayor mín. de keypoints - Mayor mín. de keypoints malos - Distancia media central, aunque tienden a separarse - Mayor máx. y mín. de keypoints totales, buenos y malos según cuadrantes - Más matches (2 de los 3 vídeos) - Mayor % de matches buenos - Menor % de matches malos - Mayor máx. y mín. de matches totales, buenos y malos según cuadrantes - Más matches entre cuadrantes diferentes - Más fotos analizadas - Menor tiempo máx. de cálculo de keypoints - Menor tiempo mín. de cálculo de keypoints - Menor tiempo medio de cálculo de keypoints - Menor tiempo mín. de cálculo de matching - Menor tiempo medio de cálculo de matching

Tabla 4-9 : Resumen de los resultados obtenidos por los extractores

4.7 Conclusiones

Tras este largo capítulo, se ha podido observar las diferentes pruebas que se han realizado para llevar a cabo este trabajo. Para ello, se han dispuesto diferentes gráficas y tablas en las que se comparaban los diferentes extractores que son objeto de estudio, analizándose en los bags disponibles. Se han tenido en cuenta 4 estudios:

- Número de keypoints
- Asociación de features
- Análisis según las condiciones de contraste
- Costes computacionales

Para cada uno de los estudios, se han realizado diferentes pruebas que se han complementado con su correspondiente gráfica o tabla en algunas ocasiones. Además, una vez expuesto el resultado, se realiza un pequeño análisis de los datos más relevantes.

Para finalizar el capítulo, se ha añadido una tabla resumen de todo lo ocurrido en el capítulo. En ella, se muestran las diferentes tecnologías junto con algunos puntos de interés y conclusiones a resaltar tras las diferentes pruebas. En el siguiente capítulo, se pasa a clarificar todo lo acontecido en estas páginas, concluyéndose con aspectos de mejora y posibles desarrollos futuros.

5 CONCLUSIONES Y DESARROLLOS FUTUROS

5.1 Conclusiones

En este capítulo final se pasa a clarificar todo lo acontecido en las páginas anteriores, en las cuales se mostraron diferentes gráficas y tablas que recogían los resultados tras analizar en diferentes situaciones a los extractores estudiados (SIFT, SURF y ORB). El objetivo de este trabajo es analizar las tecnologías existentes para conocer cual se comporta mejor: quién obtiene más keypoints, cuál consigue enlazar más puntos de manera correcta, cuál es más rápida y eficaz, estudiar el comportamiento según la luz del escenario...

Los estudios que se han realizado han sido: número de keypoints, asociación de features, estudio según las condiciones de contraste y costes computacionales. Se pasa a resumir lo ocurrido en el capítulo anterior para cada uno de los análisis, exponiendo cuál es el objetivo con cada estudio y los resultados obtenidos.

1. Número de keypoints

En primer lugar, se pasa a resumir el número de keypoints obtenido por cada uno de los extractores. Con este análisis se busca conseguir una tecnología que sea capaz de obtener mucha información de la foto y que además esta sea útil. Para ello, se define el concepto de keypoints bueno y malo, o lo que es lo mismo, que se es capaz de encontrar su análogo en otra foto o no. Si se observa la figura 4-3, se concluye que con ORB se consigue calcular más keypoints, a pesar de tener puntos repetidos. De todas formas, también se analizó el porcentaje de repetición de puntos clave en la tabla 4-1 y se aprecia que ORB apenas alcanza el 1% de repetición. Observando dicha tabla, se podría optar por elegir SURF ya que no tiene repetición de puntos, pero a pesar de ello no obtiene tanta información de los videos en comparación con ORB (figura 4-3).

Otro aspecto que destacar en el estudio de número de keypoints, es la capacidad de calcular puntos clave que hagan una buena asociación. Para ello, se recurre a la tabla 4-3 donde se muestran los porcentajes de keypoints buenos que es capaz de calcular cada extractor, donde se concluye de nuevo que ORB obtiene un alto porcentaje. Una tabla análoga a la anterior es la 4-5, que esta vez muestra los porcentajes de keypoints malos; se desea un extractor con un bajo resultado que se obtiene para ORB.

Otros análisis dentro del bloque asociado al número de keypoints es el cálculo de máximos y mínimos de puntos que se consiguen con cada tecnología. La primera gráfica que se muestra (figura 4-10), recoge el número máximo de keypoints que recoge cada extractor. Con ella, se pudo concluir que cada tecnología funciona diferente, ya que en la programación se han usado las mismas funciones que trabajan con una variable común; este valor se establece igual para los 3 extractores, siendo igual a 400. Para SIFT este valor es un límite que nunca llega a alcanzar, con ORB es un tope y con SURF se llega a sobrepasar. Cuanto más grande sea este umbral, se obtienen menos puntos clave y mayor repetición de ellos, por lo que se decide optar por el valor estándar que usa por defecto la función.

Dentro del análisis de máximos y mínimos, también se concluye un nuevo comportamiento, esta vez relacionado con los videos analizados. En concreto, en el bag grabado en la pista de baloncesto se llega a obtener mínimos iguales a cero, debido a que este vídeo se caracteriza por tener algunas fotos monocromáticas que no aportan información: dichos resultados se pueden apreciar en las figuras 4-13, 4-15 y 4-16.

Para conocer la distribución de los puntos dentro de las fotos, se recurre al cálculo de la distancia entre ellos, tanto la máxima, la mínima y la media; esta última dará una información más precisa. En el cálculo

de distancias máximas (figura 4-17), se pudo observar que cada extractor obtiene un valor máximo que se repite aproximadamente para cada vídeo, es decir, cada tecnología obtiene un resultado muy parecido para el resto de los bags. En cuanto a la gráfica relacionada con la distancia mínima (figura 4-18), se llega a la conclusión de que con ORB se obtienen valores muy pequeños, en concreto en la cancha de baloncesto, que como se comentó anteriormente está constituida por algunas fotos que no aportan información. Para el análisis de distancias medias, se mostraron diferentes gráficas (desde la figura 4-21 hasta la 4-29) que recogían la máxima, mínima y medias de todas las distancias medias. En estos análisis se contó además con la ayuda de la mediana para cada una de las figuras, para así poder analizar mejor el conjunto de datos. Todos las gráficas son muy parecidas entre sí, obteniéndose distancias preferiblemente centrales en el vídeo del puente y la pista de karting. En el vídeo de la cancha de baloncesto, todas las tecnologías coinciden en que las distancias se acercan al máximo, por lo que los puntos que se calculan están más separados; de nuevo se recuerda que este bag se caracteriza por tener planos con poca información.

El último análisis dentro de este apartado es el estudio de la localización de los puntos clave, según el cuadrante en el que se encuentre. Para ello, se ha dividido el vídeo en 12 cuadrantes, tal y como se mostró en la imagen 3-10. En todas las gráficas recogidas en este análisis (desde la figura 4-30 hasta la 4-38), se puede concluir que ORB obtiene los valores más altos, cosa que ya se sabía ya que es el extractor que más keypoints calcula. Además, cabe destacar que el lugar donde ocurren los mínimos, indiferentemente de la tecnología y del vídeo, se encuentra en una esquina. Por último, hay que añadir que en el vídeo de la pista de karts se obtienen resultados no esperados en comparación con los otros bags. Para los otros vídeos, normalmente coincidían entre sí en la localización de máximos y mínimos, o sus resultados estaban muy próximos, cosa que no ocurre en la pista de karting. Otra característica curiosa de este bag se daba cuando se analizaba el número de keypoints total, donde ORB obtenía la primera posición con diferencia excepto en el vídeo en cuestión. Para entender un poco mejor este vídeo, a partir de la página 14 se analiza dicho bag. Una de las características que destacar es que tiene una larga duración, pero a pesar de ello el número de fotos que se contabilizan no es tan elevado, por lo que los fotogramas por segundos (o fps del inglés *frames per second*) son muy bajos. Esto conlleva que el tiempo entre fotos sea más elevado, y por tanto el movimiento de los objetos resulta más brusco, lo que da a lugar que los extractores no funcionen de la forma deseada.

2. Asociación de features

En este estudio se busca la localización de un punto en la foto tomada instantes después, concepto que se conoce como match. Para llevar a cabo el enlace entre puntos, las funciones trabajan de tal forma que siempre buscan la pareja a los puntos que han calculado, por tanto los keypoints obtenidos en la última foto no obtienen un matching. Además, cuando se define si un keypoint es correcto o no está ligado a saber si el match también lo es. Por tanto, las gráficas y resultados obtenidos para este apartado son muy similares a las que se acaban de explicar en el subapartado anterior.

En primer lugar, se analizó el número de matches, que será igual al número total de keypoints restando los calculados en la última foto. Es por ello que las gráficas 4-3 y 4-39, que corresponden al número total de keypoints y de matches respectivamente, son prácticamente idénticas. Se concluye por tanto que ORB es el extractor que realiza más enlaces. Además como se acaba de comentar, para el cálculo de matches buenos (tabla 4-6) y malos (tabla 4-7) ocurre lo mismo, por lo que las conclusiones se repiten: con ORB se obtienen más matches, y un alto porcentaje de ellos (en torno al 60%) son correctos.

En este subapartado se debería analizar, al igual que antes, el número máximo y mínimo de matches encontrados, tanto total o incluso buenos y malos. Debido a lo comentado al principio de este subapartado, dichos resultados se han obtenido con anterioridad, cuando se estudió para el caso de número de keypoints.

El último análisis que se realiza en esta sección es el relacionado con la localización de los matches según el cuadrante. Para este cálculo, se ha localizado la posición de los dos puntos que realizan el match, y si los cuadrantes son diferentes se tendrá en cuenta en una variable diferente. Con la tecnología ORB, como era de esperar, una alta probabilidad de encontrar matches entre cuadrantes diferentes, debido a que es el extractor con el que se consiguen calcular más asociación de features. Para todos los

extractores, en la mayoría de los videos se consigue que alguna de las esquinas sea el lugar donde se realicen menos matches, y algún borde coincide con el lugar que se obtienen más. De nuevo, ocurre que en la pista de karting los resultados obtenidos difieren con lo acontecido en los otros vídeos, como se ha comentado en el subapartado anterior.

3. Análisis según las condiciones de contraste

Dentro de este estudio se realizaron dos pruebas: conocer el contraste de los cuadrantes y analizar la aparición o no de puntos clave en función de las condiciones lumínicas. En el primero de ellos, se comparan las gráficas resultantes de calcular el contraste en los cuadrantes con el resultado obtenido del vídeo completo. Con el video del puente, cuyas gráficas eran la 3-3 y 4-51, se pudo observar que los cuadrantes aportan mucha oscuridad, lo que se comprueba obteniéndose figuras diferentes. En el caso de la cancha de baloncesto (figura 3-6 y 4-52), se consigue llegar a que el vídeo se encuentra subexpuesto, aunque analizando cuadrantes de nuevo se consigue un fuerte pico en la parte izquierda, es decir, aportan muchas tonalidades oscuras que no se obtenían analizando la figura 3-6, que pertenece al vídeo completo. Por último, con el vídeo karts sí que se consigue una gran similitud entre las imágenes 3-9 y 4-56, concluyéndose que en esta ocasión analizar el contraste en los cuadrantes no difiere de su gráfica completa, obteniéndose un vídeo subexpuesto.

El otro análisis realizado en este subapartado consiste en calcular el número de keypoints que se obtienen según el contraste que ya se conoce en el cuadrante. Lo que se persigue es conocer cómo funciona el cálculo de keypoints, es decir, bajo qué condiciones lumínicas es más fácil encontrarlos; las gráficas son las 4-54, 4-55 y 4-56, que pertenecen al puente, baloncesto y karts respetivamente. En todas las gráficas se pudo concluir que en cuanto se obtienen contrastes altos o claros, la capacidad de cálculo de keypoints es prácticamente nula. Esta idea está contrastada con el alto número de puntos clave que se observan en la parte izquierda de la gráfica, que pertenecen a imágenes oscuras. Concluyendo, se obtienen más puntos clave en escenarios con tonos oscuros y semioscuros.

4. Costes computacionales

En este subapartado se busca conocer que extractor funciona de manera más eficaz y rápida. Se han realizado tres mediciones: tiempo de cálculo de keypoints, tiempo de cálculo de matches y número de fotos analizadas. Se pretende obtener una tecnología que obtenga bajos tiempos y sea capaz de analizar una gran parte de las fotos que componen al vídeo.

Analizando las gráficas 4-58 y 4-59, que muestran los tiempos máximos y mínimos de cálculo de puntos clave, se observa que ORB es con diferencia el extractor más rápido. A pesar de ello, también se cuenta con la ayuda de las gráficas que recogen los tiempos medios de cálculo, que van desde la figura 4-60 hasta la 4-68. La mayoría de las gráficas obtienen medias cercanas al mínimo, aunque con ORB esta tendencia es más clara. En las gráficas de este extractor se observa como la curva obtiene un pico muy alto al comienzo de la figura, y que después de él apenas se obtienen datos. Esto es debido a que con este extractor, los primeros tiempos de cálculo de las primeras fotos son un poco elevados, pero al analizar muchas fotos y obtener tiempos bajos, los valores altos apenas son apreciables.

El otro análisis temporal que se ha realizado es el perteneciente a la asociación de features. Si se observan los valores en comparación con los relativos a cálculo de keypoints, estos últimos obtienen números más altos, es decir, se tarda más en calcular puntos clave que en buscar su match. Las gráficas pertenecientes a los tiempos de matches van desde la figura 4-69 hasta la 4-79. Las dos primeras muestran los tiempos máximo y mínimo respectivamente, obteniéndose el mayor máximo para SIFT en dos videos excepto en el puente, donde la primera posición es ocupada por ORB. La figura siguiente muestra los tiempos mínimos, destacándose el hecho de que en la pista de baloncesto los tiempos son muy pequeños en comparación con el resto. Debido a que con los tiempos máximos y mínimos no se tiene una idea clara, también se mostraron diferentes figuras (desde 4-71 hasta 4-79) que recogen los tiempos medios. De nuevo ocurre la situación comentada con los keypoints, es decir, a pesar de que con SURF y SIFT los tiempos medios están en la mitad izquierda, con ORB esta cercanía es mayor incluso se nota un fuerte pico en algunas ocasiones.

Por último en el análisis de costes computacionales, se muestra la tabla 4-8 que recoge el porcentaje de

fotos analizadas por cada extractor respecto al total que componen cada vídeo. Dicha información se puede ver a partir de la página 11, donde se analizan las características de cada bag. En la tabla se puede apreciar que ORB es con diferencia la tecnología que más fotos llega a analizar. También se destaca que en el vídeo de la pista de karts, todos los extractores obtienen cifras muy similares, llegando a analizar casi en su totalidad al vídeo en cuestión. Como ya se comentó anteriormente, este vídeo tiene una larga duración y pocas fotos, por lo que en cada cambio de imagen a cada extractor le ha dado tiempo de poder analizar sus características.

Con todo lo comentado anteriormente, se puede concluir que a lo largo de los años y las investigaciones, se ha podido ir mejorando las diferentes técnicas en la extracción visual. La primera de ellas (SIFT), como se ha podido observar da a lugar a muchas repeticiones de puntos, cosa no deseable, además que los porcentajes y tiempos no son los deseados. Después aparece SURF, pero no es hasta la llegada de ORB que se llegan a tener los resultados y eficacia que se estaba persiguiendo años atrás. Con esta tecnología se puede obtener unos tiempos de cálculo menores, altos porcentajes de análisis de las fotos y una gran cantidad de enlaces correctos entre los puntos. A pesar de todo, existen muchos aspectos que mejorar, como el mejor funcionamiento en condiciones lumínicas claras o aumentar el porcentaje de matches correctos. Como se puede ver, los cambios y perfeccionamientos aparecen con los años, por lo que existe una alta probabilidad de que aparezca una nueva tecnología que supere a sus antecesoras.

5.2 Desarrollos futuros

Uno de los principales puntos fuertes de este trabajo es que existen muchos más extractores que estudiar. Con ellos, se podría conocer su comportamiento y compararlos con los que aquí se han mostrado. Dentro de una de la librerías de OpenCV [16], se pueden observar los diferentes detectores de keypoints que están programados y desarrollados, por ejemplo, KAZE, BRISK, AKAZE entre otros.

Al ser una programación trabajada y hecha por experimentación, existen muchos factores que mejorar en el código, no sólo la limpieza de este. Una de las primeras cosas que se puede hacer es trabajar previamente con los vídeos: eliminar el posible ruido, usar mejor tecnología y mejorar así la resolución ... Al haber concluido que con la claridad los extractores no son tan eficientes, se puede estudiar de nuevo en fondos más claros, con ayuda de focos y equipo especial para sacar mejores conclusiones.

Con relación a los bags que se han estudiado, se podría contar con más material, es decir, tener más vídeos que complementen el trabajo. Además, debido a que las cámaras usadas son estéreas, o lo que es lo mismo, se tiene una cámara izquierda y otra derecha, las posibilidades de estudio se amplían. Con ello, se puede estudiar todo lo que se ha presentado en este trabajo, ampliando los cálculos a la otra cámara, o incluso comparando los keypoints que han analizado cada una de ellas y ver el desplazamiento de los objetos a lo largo del escenario.

A raíz de conocer la localización de los keypoints, el siguiente paso sería reconocer formas. Por ejemplo, en el vídeo de la cancha de baloncesto se pueden observar las canastas, por lo que se puede identificarlas y ver su desplazamiento en el recorrido del dron. Este estudio de reconocimiento amplía las posibilidades, ya que en la actualidad es un ámbito que se busca desarrollar: reconocimiento de caras en videovigilancia, búsqueda de objetos en un escenario ...

REFERENCIAS

- [1] Edwin Wise, *Applied Robotics*, 1999.
- [2] Hugh Durrant-Whyte, Fellow, IEE and Tim Bailey, *Simultaneous Localisation and Mapping (SLAM): Part I The Essential Algorithms*, 2006.
- [3] https://canal.uned.es/uploads/material/Video/49999/Presentaci_n_Oscar_Reinoso.pdf
- [4] Søren Riisgaard and Morten Rufus Blas, *SLAM for Dummies*, 2004.
- [5] David G. Lowe, *Distinctive Image Features from Scale-Invariant Keypoints*, 2004.
- [6] <https://opencv.org/>
- [7] <http://aishack.in/tutorials/sift-scale-invariant-feature-transform-keypoint-orientation/>
- [8] <https://aishack.in/tutorials/sift-scale-invariant-feature-transform-features/>
- [9] Herbert Bay, Tinne Tuytelaars and Luc Van Gool, *SURF: Speeded Up Robust Features*, 2006.
- [10] https://docs.opencv.org/3.0-beta/doc/py_tutorials/py_feature2d/py_surf_intro/py_surf_intro.html#surf
- [11] Rublee, Ethan & Rabaud, Vincent & Konolige, Kurt & Brandyk and Gary, *ORB: an efficient alternative to SIFT or SURF*, 2011.
- [12] <http://iie.fing.edu.uy/investigacion/grupos/gti/timag/trabajos/2011/keypoints/FundamentoRANSAC.pdf>
- [13] https://scikit-learn.org/stable/auto_examples/linear_model/plot_ransac.html
- [14] <https://www.ros.org/>
- [15] Katherine L. Monti, *Folded Empirical Distribution Function Curves-Mountain Plots*, *The American Statistician* Vol.49 No. 4, 1995.
- [16] https://docs.opencv.org/3.4/d5/d0d/features2d_8hpp.html

