

07-011

DEVELOPMENT OF AN APPLICATION FOR OBJECT RETRIEVAL AND EVALUATION OF THE LOCAL INVARIANT DESCRIPTORS USED

Fernández-Robles, Laura; Alfonso-Cendón, Javier; Castejón, Manuel; García-
Olalla, Óscar; Alegre, Enrique
Universidad de León

Retrieval of objects in big datasets of images and videos is one of the tasks of the European Project ASASEC (Advisory System Against Sexual Exploitation of Children). Spanish National Police Department of Technological Information is interested in selecting a box around an object considered as evidence and finding the same object in a great collection of images and videos registered as evidences of crimes of sexual exploitation of children. For this purpose, an application that allows the user to choose one image, select a box over the image using the mouse and look for similarities between the region of interest and the rest of the images in a folder has been developed. In order to describe the region of interest and the rest of the images, several well-known robust local invariant descriptors have been used such as: SURF, SIFT and COSFIRE. Experiments were performed on a 100 images dataset taken from videos recorded in the same room, Lonelygirl dataset, with different poses of the camera, occlusions and illumination changes. Results show the benefit of using SIFT and COSFIRE opposite to SURF and HOG.

Keywords: *ASASEC; object retrieval; local invariant descriptors*

DESARROLLO DE UNA APLICACIÓN DE RECUPERACIÓN DE OBJETOS Y ANÁLISIS DE LOS DESCRIPTORES LOCALES INVARIANTES UTILIZADOS

Una de las tareas del proyecto europeo ASASEC (Advisory System Against Sexual Exploitation of Children) es la recuperación de objetos en grandes bases de imágenes y vídeos. La Brigada de Información Tecnológica de la Policía Nacional española está interesada en seleccionar un recuadro alrededor de un objeto que consideren una evidencia y localizar ese mismo objeto en una gran colección de imágenes y vídeos de delitos de explotación sexual de niños. Con este fin, se ha desarrollado una aplicación que permite al usuario seleccionar una imagen, seleccionar un recuadro sobre la imagen con el ratón y buscar similitudes a la región de interés marcada en el resto de imágenes de una carpeta indicada. A la hora de describir la región de interés y el resto de imágenes se han utilizado varios descriptores robustos basados en características locales de la imagen como son los conocidos métodos: SURF, SIFT y COSFIRE. Los experimentos fueron evaluados en una base de 100 imágenes tomadas de vídeos grabados en una misma habitación, base de imágenes Lonelygirl, con diferentes posiciones de la cámara, occlusiones y cambios en iluminación. Los resultados muestran la conveniencia de utilizar SIFT y COSFIRE frente al resto para resolver este problema.

Palabras clave: *ASASEC; recuperación objetos; descriptores locales invariantes*

Correspondencia: Laura Fernández Robles l.fernandez@unileon.es

Agradecimientos: Proyecto nacional DPI2012- 36166 del Gobierno de España. Beca FPU AP2010- 0947 del Gobierno de España.

1. Introducción

La Universidad de León está participando en el proyecto europeo ASASEC (Advisory System Against Sexual Exploitation of Children) en conjunto con diversas entidades públicas y privadas. La policía posee una gran recopilación de imágenes y vídeos de delitos de explotación infantil. Uno de los objetivos del proyecto es la búsqueda automática de un objeto dado, que la policía reconozca de pruebas anteriores o considere interesante, en esa colección de imágenes y vídeos. La gran variedad de escenarios, posiciones de cámara, calidades de vídeo, iluminaciones, etc. que dicha colección posee supone un gran reto que no se ha conseguido solventar en la literatura.

Por tanto, idealmente el método aplicado debe ser invariante a cambios en escala, rotación, iluminación y ruido además de ser capaz de detectar el objeto de manera local en condiciones de oclusión. Los métodos basados en características locales invariantes se han posicionado entre los más utilizados para estos fines.

En 2004, Lowe creó el método scale invariant feature transform (SIFT) para extraer características locales distintivas de imágenes en escala de grises pudiendo ser utilizado para reconocer la misma característica en imágenes donde el objeto presenta diferentes poses. SIFT es invariante a rotaciones y escalado y parcialmente invariante frente a cambios en iluminación y transformaciones afines. Además, al ser un método local, permite reconocer el objeto de búsqueda cuando se encuentra parcialmente ocluido. Estas características han hecho que SIFT se convierta en uno de los algoritmos más utilizados en visión artificial para el reconocimiento de objetos (Tian et al., 2014). Nuevas investigaciones van encaminadas a la utilización de SIFT en tiempo real para la detección de objetos (Piccinini, Patri y Cucchiara, 2012) y en la mejora de la eficiencia del algoritmo (Yu et al., 2013) (Liao, Liu y Hui, 2013).

Speeded-up robust feature (SURF) (Bay, Tuytelaars y Van Gool, 2008) es un algoritmo basado en SIFT que permite aumentar la velocidad de cálculo manteniendo o incluso superando la eficacia de SIFT según sus autores. Las principales diferencias son el uso de la imagen integral que permite cambiar el tamaño del filtro a aplicar en lugar de utilizar la pirámide de gaussianas del método SIFT. SURF se sigue utilizando en la actualidad para la recuperación de objetos (Yang et al., 2014).

Debido a la eficiencia de SIFT y SURF, recientemente muchos trabajos han ido en la línea de mejorar su tiempos de cálculos para poder usarlos en grandes bases de datos. Özuysal et al. (2010) presentaban un método de rápido reconocimiento de puntos clave utilizando Ferns, lo que evita el cálculo costoso del preprocesamiento de parches mediante el uso de cientos de características binarias. Otros trabajos que siguen esta línea son los conocidos binary robust independent elementary feature (BRIEF) (Calonder et al., 2010), oriented fast and rotated BRIEF (ORB) (Rublee et al., 2011), binary robust invariant scalable keypoints (BRISK) (Leutenegger, Chli, and Siegwart, 2011) y fast retina keypoint (FREAK) (Alahi, Ortiz and Vanderghyest, 2012). Su principal contribución es que las cadenas binarias obtenidas mediante una simple comparación de pared de imágenes de intensidad puede describir eficientemente un punto clave, por ejemplo una región de una imagen.

Dalal y Triggs (2005) propusieron un nuevo algoritmo de descripción de características basado en histogramas de gradientes orientados (HOG – histogram of oriented gradients) normalizados en áreas locales de una imagen. Aunque fue inicialmente propuesto para la detección de personas, este método se ha extendido posteriormente para la detección de diferentes clases de objetos con resultados empíricamente buenos. La idea fundamental de HOG es que la apariencia y forma local de un objeto pueden ser descritas mediante la distribución de gradientes de intensidad locales o direcciones de bordes, incluso cuando el

conocimiento de las posiciones de dichos gradientes o bordes no es precisa. Hu y Collomosse (2013) han trabajado en la recuperación de imágenes utilizando HOG y más recientemente en 2015, Liang y Juang en la clasificación de objetos.

En 2013, Azzopardi y Petkov presentaron un método que permite configurar filtros entrenables, denominados filtros COSFIRE, para el reconocimiento de patrones. Un filtro COSFIRE se puede entrenar automáticamente para ser selectivo para un patrón de contorno local de un objeto determinado mediante una única imagen de dicho objeto. La configuración implica la selección de ciertos canales de un banco de filtros Gabor y la determinación de los valores de parámetros de emborronamiento y desplazamiento. La respuesta de un filtro COSFIRE se calcula como la media ponderada de las respuestas emborronadas y desplazadas de los filtros Gabor seleccionados. COSFIRE ha probado ser eficiente en la clasificación de objetos para tareas de reconocimiento de señales (Azzopardi y Petkov, 2013) y para el reconocimiento de objetos realizado por un robot doméstico de recogida de objetos (Azzopardi y Petkov, 2014).

En este trabajo vamos a comparar los métodos SIFT y SURF, por ser los más extendidos, y el método COSFIRE para comprobar si los buenos resultados de las aplicaciones previas pueden ser aplicables a nuestro problema.

El resto del artículo se organiza de la siguiente forma. La Sección 2 describe los objetivos del trabajo. En la Sección 3, se explica el método desarrollado. La Sección 4 desarrolla la evaluación del método en la base de datos propuesta y presenta los resultados alcanzados. Por último, se describen las conclusiones obtenidas.

2. Objetivo

El principal objetivo de este trabajo es el estudio de metodologías para el desarrollo de un software que permita la recuperación de objetos en grandes bases de imágenes y vídeos.

3. Metodología

3.1 Introducción al proyecto

ASASEC es un proyecto cofinanciado por la Unión Europea (UE) dentro del programa "Prevención y Lucha Contra el Crimen Se abre en nueva ventana", que es parte del programa "Seguridad y Defensa de las Libertades".

La principal meta del proyecto ASASEC (Advisory System Against Sexual Exploitation of Children) es el desarrollo de una solución tecnológica para mejorar las técnicas actuales en la lucha contra la pornografía infantil. El usuario final en este caso es la Brigada de Investigación Tecnológica de la Dirección General de Policía, aunque se colabora con otros cuerpos y fuerzas de seguridad a nivel internacional.

El proyecto está planificado con una duración estimada de tres años y su alcance pretende cubrir los aspectos de:

- Gestión de casos y evidencias.
- Análisis forense de evidencias.
- Tratamiento y explotación de datos obtenidos de casos y evidencias.

Dentro de este proyecto, una parte se centra en el diseño, desarrollo y evaluación de la aplicación de técnicas de visión artificial para analizar evidencias con contenido de pornografía infantil.

Más información sobre el proyecto puede encontrarse en su página web¹.

3.2 Ingeniería del software

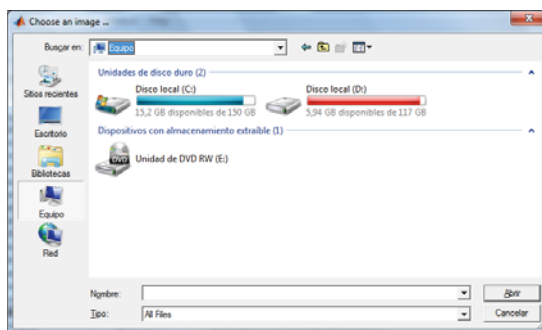
Este proyecto se ha desarrollado mediante la metodología de ingeniería del software conocida como Proceso Unificado Ágil o por su nombre y siglas en inglés AUP (Agile Unified Process) de Scott Ambler (2014). De las cuatro fases de esta metodología, iniciación, elaboración, construcción y transición, nos encontramos en la etapa de construcción. El objetivo de esta fase consiste en construir software de manera incremental según las prioridades de los usuarios. El trabajo expuesto en este artículo se presenta como parte de uno de los primeros entregables, que se irán liberando de manera incremental con el tiempo. El prototipado es por tanto la técnica fundamental sobre la que se sustenta este enfoque. En cada entrega, el cliente realiza un análisis del prototipo que se le entrega y elabora una lista de prioridades que se han de cumplir para el siguiente entregable. En nuestro caso, se ha utilizado un tiempo de 1 semana entre cada una de las versiones entre la Universidad de León y el INCIBE y de 1 mes entre la Universidad de León y la Brigada de Investigación Tecnológica de la Dirección General de Policía.

La continua realimentación obtenida con este tipo de desarrollo facilita a los programadores su labor para la elaboración de un software orientado a satisfacer las necesidades del proyecto sin desviarse de los intereses del usuario final. Esto hace que se eviten retrasos en el entregable final o que se produzca una corrupción del alcance del proyecto, y por tanto, el coste final del proyecto se encontrará más cercano al coste inicial estimado.

3.3 Interfaz

En primer lugar, se le pide al usuario que elija una imagen del directorio que contenga el objeto de interés. A continuación, el usuario debe marcar una región de interés rectangular que contenga dicho objeto. Este proceso se muestra en la Figura 1.

Figura 1: (a) Selección de una imagen de un directorio. (b) Selección manual de una región de interés marcada en azul sobre la imagen previamente seleccionada.



(a)



(b)

3.4 Descripción de la región de interés

i) SIFT

En la Figura 2a se muestra una representación de los 29 descriptores para cada punto clave detectado con SIFT. El origen de cada flecha se localiza en la posición del punto clave detectado. La longitud de la flecha es proporcional a la escala a la que se encontró el punto

¹ <http://www.asasec.eu>

clave. Y su orientación refleja la orientación del punto clave obtenida a partir del histograma de orientaciones de gradiente. Se ha seguido la configuración por defecto propuesta por Lowe y su código (Lowe,2005).

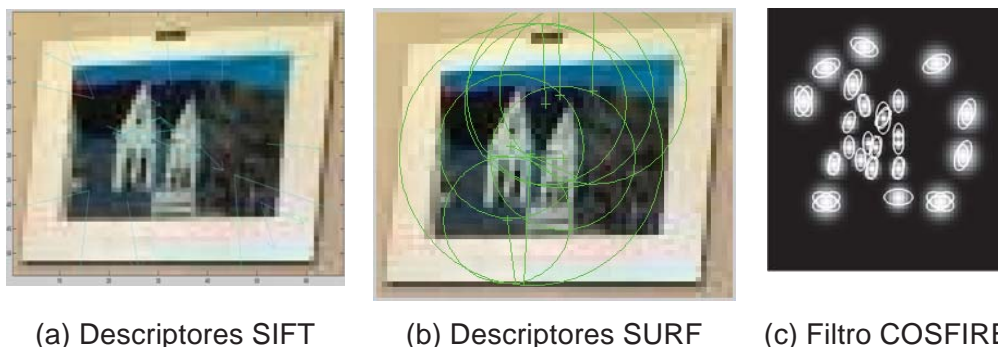
ii) SURF

La Figura 2b muestra una representación de los 7 descriptores SURF detectados. A diferencia de SIFT, en cada punto clave solamente es posible definir un descriptor. Los centros de cada circunferencia, marcado con una cruz, indica la localización de los puntos clave. Los radios de las circunferencias son proporcionales a la escala en la que se localizó el punto clave. El segmento radial que une el centro con la circunferencia se dibuja con la orientación del punto clave detectado. SURF se ha aplicado con 3 octavas y 4 niveles de escala por octava.

iii) COSFIRE

La Figura 2c muestra la estructura del filtro COSFIRE configurado. Las elipses representan los soportes del detector de bordes configurados. Además, las elipses están representadas en relación a las longitudes de onda y las orientaciones de los filtros Gabor que dan lugar a ellas. Este filtro ha sido entrenado para detectar arreglos espaciales locales con representaciones muy similares a las que configuran cada elipse de este patrón. Las manchas blancas son los mapas de intensidad de las funciones gaussianas utilizadas para el emborronamiento de las respuestas de los filtros Gabor. La configuración del filtro COSFIRE se ha realizado con los siguientes valores para los parámetros: $\rho=\{0,4,9,15,22\}$; $t_1=0.05$; $t_2=0.3$; $t_3=0.99$; $\sigma_0=0.67$; $\alpha=0.1$; $\text{mintupleweight}=0.5$; $\text{psilist}=0$ (invarianza a rotación); $\text{upsilon list}=\{0.75,1\}$ (invarianza a escala); $\text{reflection}=0$ (invarianza a reflexión); $\text{thetalist}=\{\pi/8, i=0\dots 8\}$; $\text{lambdalist}=\{4,4\sqrt{2}\}$; $\text{phaseoffset}=\pi/2$; $\text{halfwaverect}=0$; $\text{bandwidth}=1.5$; $\text{aspectratio}=0.5$; sin inhibición.

Figura 2: Descripción de la región de interés con los diferentes métodos evaluados



3.5 Detección del objeto en cola en las imágenes del data set

i) SIFT

Del mismo modo que describíamos la región de interés, ahora describimos la imagen de búsqueda. La distancia entre pares de descriptores de la región de interés y de la imagen de búsqueda se ha calculado como el ratio de sus ángulos, es decir, como la inversa del coseno del producto escalar de los vectores normalizados (Lowe, 2005). Sólo se consideran válidos los emparejamientos de pares de descriptores cuyo ratio entre el ángulo obtenido por el par de descriptores más cercano respecto del ángulo del segundo par de descriptores más cercano sea menor que 0.6 El valor de la inversa del coseno del producto escalar de los vectores normalizados del par de descriptores más cercano constituye el valor de fuerza

para la detección del objeto en la imagen de búsqueda. La posición del objeto detectado vendrá dada por las coordenadas del punto clave cuyo descriptor se encuentra a la distancia más cercana.

ii) SURF

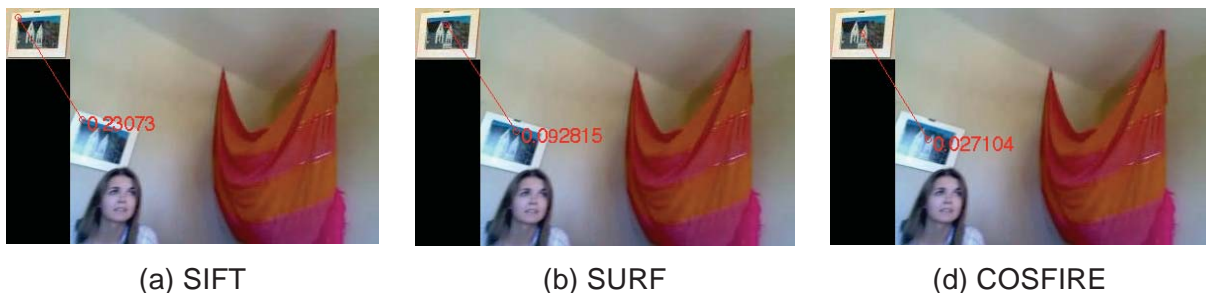
Utilizamos la distancia Euclídea cuadrada para medir la distancia entre pares de descriptores. Los descriptores son normalizados a vectores unitarios por lo que el rango de distancias entre dos vectores varía entre 0 (emparejamiento perfecto) y 4. Rechazamos emparejamientos cuya distancia sea mayor de 2. Si el ratio entre la distancia de los primeros vecinos más cercanos a los segundos vecinos más cercanos es mayor de 0.6, consideramos que el emparejamiento es ambiguo y lo eliminamos. La distancia del par de descriptores con menor distancia indica la fuerza de la detección del objeto en la imagen de búsqueda. La posición, al igual que en SIFT, se determina a partir de las coordenadas del punto clave cuyo descriptor produce el emparejamiento más fuerte.

iii) COSFIRE

El filtro COSFIRE configurado se aplica a la imagen de búsqueda tal y como describen sus autores (Azzopardi y Petkov, 2013). Los parámetros utilizados son los mismos que en la configuración. La fuerza de la detección del objeto en la imagen de búsqueda se define como el mayor valor de la respuesta del filtro COSFIRE aplicado en cada píxel de dicha imagen. La localización viene dada por las coordenadas del píxel con mayor valor de respuesta.

La Figura 3 muestra un ejemplo de la detección de la región de interés mostrada en la Figura 2 en una de las imágenes del data set Lonelygirl. A pesar de que el cuadro aparece en esta imagen con diferente pose e iluminación, todos los métodos evaluados consiguen detectarlo correctamente.

Figura 3: Detección de la región de interés de la Figura 2 con la imagen 21 del data set. Los valores numéricos indican las fuerzas de la detección en cada caso. Los círculos rojos muestran las posiciones del píxel que ha obtenido el máximo valor de fuerza.



4. Resultados

4.1 Data set

Se han extraído 100 imágenes de 320x240 píxeles tomadas de vídeos de la serie de YouTube "lonelygirl15 Season One" (lonelygirl15, 2014). Las imágenes se han extraído a partir de 50 vídeos que se han grabado en una misma habitación en diferentes días, con diferentes posiciones de la cámara y cambios en iluminación. Los 100 fotogramas de los vídeos se han elegido de manera que haya una amplia variedad en las posiciones de captura, iluminación, enfoque, etc. Por tanto este data set se aproxima a un vídeo casero realizado con una cámara web en la que las personas presentes pueden ocluir objetos del

entorno y las condiciones de captura son variables, como es el caso de la aplicación de nuestro estudio.

Para automatizar la experimentación se ha creado un ground truth de las imágenes, indicando las coordenadas del punto medio de la región de los objetos evaluados.

Tanto el data set, como los ground truth para cada objeto están disponibles². Denominamos a este data set Lonelygirl.

4.2 Experimentación y resultados

Hemos aplicado el método explicado en la Sección 2 al data set Lonelygirl, obteniendo los valores de fuerza y las coordenadas del píxel de respuesta para cada método y para cada una de las 100 imágenes.

Se considera que un objeto detectado está bien localizado si se encuentra dentro de un recuadro de tamaño igual al tamaño de la región de interés más 10 píxeles en cada dimensión y centrado alrededor del punto definido en el ground truth. Se considera como verdadero positivo (VP) a una imagen que contiene el objeto y éste es localizado dentro de los márgenes del ground truth. Se considera un falso positivo (FP) cuando el objeto en cola no está presente en la imagen pero éste es detectado y también cuando el objeto está presente y es detectado pero fuera de los márgenes del ground truth. Por último, un falso negativo (FN) se obtiene cuando el objeto está presente en la imagen pero no es detectado y también cuando el objeto está presente y es detectado pero fuera de los márgenes del ground truth. Además, consideramos que el objeto es detectado cuando el valor de fuerza es superior a un valor de umbral determinado. La lista de imágenes con objetos detectados ordenada de mayor a menor por su valor de fuerza constituye lo que en la recuperación de objetos se denomina una hit list. Hemos tomado unos valores de umbral que hacen posible ir descartando cada imagen devuelta en la hit list de una en una.

Finalmente, se calculan los valores de precisión, recall y F-Score para diferentes umbrales y se elige el umbral que provee un mayor F-Score.

$$Precision = VP / (VP + FP) \quad (1)$$

$$Recall = VP / (VP + FN) \quad (2)$$

$$F-Score = 2 \cdot (Recall \cdot Precision) / (Recall + Precision) \quad (3)$$

Los experimentos se evaluaron con dos regiones de interés diferentes. En el primer caso, se obtuvieron resultados con la región de interés mostrada en la Figura 1, 2 y 3. Dichos resultados se muestran en la Tabla 1.

Tabla 1. Resultados de la recuperación del objeto mostrado en la Fig. 1.

	COSFIRE	SIFT	SURF
F-Score	0.8119	0.7800	0.7711
Precision	0.8039	0.7800	0.9697
Recall	0.8200	0.7800	0.6400
Umbral	0.040	0.326	0.077

² <http://pitia.unileon.es/varp/node/402>

La región de interés no estaba ocluida en ninguna de las imágenes del data set, o bien se encontraba completamente presente o bien no aparecía. El método COSFIRE consigue los mejores resultados en términos de F-Score, recall y precision. Es importante obtener un valor alto de recall en nuestra aplicación ya que es más favorable tener falsos positivos que falsos negativos y, por tanto, recuperar imágenes que no tienen el objeto que dejar de recuperar imágenes que contienen el objeto. Esto es así, porque tras la detección automática del objeto en cola, un policía verificará visualmente si los resultados devueltos son buenos. Fácilmente podrá descartar falsos positivos pero no podrá localizar un falso negativo en una gran base de datos.

El segundo experimento evaluó la recuperación del póster mostrado en la Figura 4. El póster se encuentra ocluido en algunas imágenes y en otras aparece incompleto en los bordes de la imagen.

Figura 4: Póster tomado como región de interés del segundo experimento. (a) Selección de una región de interés. (b) Región de interés aumentada.



Los resultados de este experimento se muestran en la Tabla 2. En este caso, SIFT obtiene los mejores resultados. El recall ha descendido en todos los casos debido a la dificultad añadida del experimento. Gracias a este experimento hemos detectado un punto débil de los filtros COSFIRE. Éstos no pueden detectar objetos cuando aparecen recortados por los bordes de la imagen ya que nunca va a existir una región que contenga todo el objeto y por tanto que pueda ser descrita por el filtro. Esto se debe a la función de media geométrica ponderada que COSFIRE utiliza por defecto para la combinación de las diferentes respuestas de filtros Gabor, en la que si una de las respuestas Gabor es nula, la respuesta del filtro COSFIRE es también nula. Sin embargo esta función no es intrínseca del método y otras como una media aritmética ponderada podrían ser evaluadas. En cambio, los métodos SURF y SIFT pueden seguir encontrando puntos de interés que se localicen y describan en la parte del objeto que permanece en la imagen.

Tabla 2. Resultados de la recuperación del objeto mostrado en la Fig. 4.

	SIFT	COSFIRE	SURF
F-Score	0.8113	0.6667	0.5176
Precision	0.9773	0.8919	0.9565
Recall	0.6935	0.5323	0.3548
Umbral	0.283	0.07	0.065

Conclusiones

Se ha evaluado la recuperación de objetos en un data set de imágenes tomadas a partir de vídeos de baja calidad en una misma habitación. El uso de este data set permite representar una situación similar pero simplificada de la recuperación de objetos en vídeos del proyecto ASASEC. COSFIRE y SIFT son los métodos más adecuados para la recuperación de objetos. Cuando los objetos se encuentran ocluidos o incompletos, SIFT consigue los mejores resultados. Este conforma un primer paso para un estudio mayor de la recuperación de objetos en grandes bases de datos en escenas de interior y vídeos de baja calidad.

Futuros trabajos pueden orientarse a la búsqueda de soluciones para que COSFIRE recupere objetos cortados por el borde de una imagen y a mejorar los resultados de COSFIRE en regiones de interés de bajo contraste como lámparas donde se ha visto que su rendimiento decrece considerablemente.

Referencias

- Alahi, A., Ortiz, R., & Vandergheynst, P. (2012). FREAK: Fast Retina Keypoint. *IEEE Conference on Computer Vision and Pattern Recognition*, (pp. 510-517). IEEE. DOI: 10.1109/CVPR.2012.6247715.
- Azzopardi, G., & Petkov, N. (2014). Ventral-stream-like shape representation: from pixel intensity values to trainable object-selective cosfire models. *Frontiers in Computational Neuroscience*. DOI: 10.3389/fncom.2014.00080.
- Azzopardi, G., & Petkov, N. (2013). Trainable COSFIRE filters for keypoint detection and pattern recognition. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 35(2), 490-503.
- Bay, H., Tuytelaars, T., & Van Gool, L. (2008). SURF: Speeded Up Robust Features. *Computer Vision and Image Understanding (CVIU)*, 110(3), 346-359.
- Calonder, M., Lepetit, V., Strecha, C., & Fua, P. (2010). Brief: Binary robust independent elementary features. *European Conference on Computer Vision (ECCV)*, (pp. 778-792).
- Dalal, N. & Triggs, B. (2005). Histograms of oriented gradients for human detection. *Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005)* (pp. 886-893). IEEE. DOI: doi: 10.1109/CVPR.2005.177.
- Hu, R., & Collomosse, J. (2013). A performance evaluation of gradient field HOG descriptor for sketch based image retrieval. *Computer Vision and Image Understanding*, 117(7), 790-806. DOI: <http://dx.doi.org/10.1016/j.cviu.2013.02.005>.
- Leutenegger, S., Chli, M., & Siegwart, R. (2011). Brisk: Binary robust invariant scalable keypoints. *IEEE International Conference on Computer Vision (ICCV)*, (pp. 2548-2555). IEEE. DOI: 10.1109/ICCV.2011.6126542.
- Liang, C.-W., & Juang, C.-F. (2015). Moving object classification using local shape and HOG features in wavelet-transformed space with hierarchical SVM classifiers. *Applied Soft Computing*, 28, 483-497. DOI: <http://dx.doi.org/10.1016/j.asoc.2014.09.051>.
- Liao, K., Liu, G., & Hui, Y. (2013). An improvement to the SIFT descriptor for image representation and matching. *Pattern Recognition Letters*, 34(11), 1211-1220. DOI: <http://dx.doi.org/10.1016/j.patrec.2013.03.021>.
- lonelygirl15 (2014, 29 Junio). *lonelygirl15 Season One* [Video files]. Videos publicado en <https://www.youtube.com/playlist?list=PLF1BED25959F378B7>

- Lowe, D. (2005, Julio). Demo Software: SIFT Keypoint Detector. Obtenido el 10 de abril de 2015, desde <http://www.cs.ubc.ca/~lowe/keypoints/>.
- Lowe, D. (2004). Distinctive image features from scale invariant keypoints. *International Journal of Computer Vision*, 2, 91–111.
- Lowe, D (1999). Object Recognition from Local Scale-Invariant Features. *Seventh Int'l Conference on Computer Vision* (pp. 1150- 1157).
- Özuysal, M., Calonder, M., Lepetit, V., & Fua, P. (2010). Fast keypoint recognition using random ferns. *Pattern Analysis and Machine Intelligence*, 32, 448-461.
- Piccinini, P., Prati, A., & Cucchiara, R. (2012). Real-time object detection and localization with SIFT-based clustering, *Image and Vision Computing*, 30(8), 573-587.
- Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. (2011). Orb: an efficient alternative to sift or surf. *International Conference on Computer Vision*, (pp. 2564-2571). IEEE. DOI: 10.1109/ICCV.2011.6126544.
- Ambler, S. (2014). The Agile Unified Process (AUP). Obtenido el 20 de mayo de 2015, desde <http://www.ambysoft.com/unifiedprocess/agileUP.html>
- Tian, X., Jiao, L., Liu, X., & Zhang, X. (2014). Feature integration of EODH and Color-SIFT: Application to image retrieval based on codebook. *Signal Processing: Image Communication*, 29(4), 530-545.
- Yang, H.-Y., Li, Y.-W., Li, W.-Y., Wang, X.-Y., & Yang, F.-Y. (2014). Content-based image retrieval using local visual attention feature. *Journal of Visual Communication and Image Representation*, 25(6), 1308-1323. DOI: <http://dx.doi.org/10.1016/j.jvcir.2014.05.003>.
- Yu, J., Qin, Z., Wan, T., & Zhang, X. (2013). Feature integration analysis of bag-of-features model for image retrieval. *Neurocomputing*, 120(23), 355-364. DOI: <http://dx.doi.org/10.1016/j.neucom.2012.08.061>.