

# Sistema robusto para la detección y seguimiento de personas en aplicaciones video-vigilancia.

Marcos Baptista Ríos, Carlos Martínez García, Cristina Losada Gutiérrez y Marta Marrón Romera

Departamento de Electrónica. Universidad de Alcalá. Alcalá de Henares, Madrid.

{marcos.baptista, carlos.martinez, losada, marta}@depeca.uah.es

**Abstract**— En este trabajo se presenta una solución robusta para la detección y seguimiento de un número variable de personas, a partir de secuencias de imágenes RGB, para aplicaciones de video-vigilancia. El sistema desarrollado permite tanto la detección de personas en imágenes de video-vigilancia, como el conteo de las mismas para diferentes aplicaciones como el control de accesos o el control de aforos. Para la detección de personas se emplea un descriptor HOG a 64 escalas diferentes y un clasificador SVM lineal. Además, se incorpora un seguidor para múltiples personas basado en un banco de filtros de Kalman que permiten incrementar la robustez del sistema frente a oclusiones o pérdidas temporales de la detección. Para la validación del algoritmo desarrollado se han realizado múltiples pruebas experimentales empleando tanto la base de datos de imágenes del proyecto CAVIAR, ampliamente utilizada por la comunidad científica, como una base de datos diseñada, grabada y etiquetada para esta aplicación. Los resultados obtenidos han permitido validar el proceso de detección, así como demostrar que la incorporación de la etapa de seguimiento incrementa la robustez frente a oclusiones o errores en la detección, con un aumento mínimo de la carga computacional.

**Index Terms**—Detección de personas, filtro de Kalman, imágenes monoculares, secuencias RGB, seguimiento de múltiples personas.

## I. INTRODUCCIÓN

DURANTE las últimas décadas, temas como el análisis de secuencias de video y la interpretación de imágenes han adquirido una gran importancia debido al potencial de sus múltiples aplicaciones [1] como la seguridad, la video-vigilancia [2], los espacios inteligentes, el apoyo a la dependencia o el marketing. Más concretamente, la detección de personas para el reconocimiento de las actividades que realizan [3] es un tema que despierta un gran interés en la comunidad científica.

Los sensores típicamente utilizados en estas aplicaciones son cámaras de color y de profundidad (múltiples en algunos casos), y las características extraídas abarcan tanto el bajo, como el medio y alto nivel.

Existen numerosos trabajos en la literatura científica cuyo objetivo es la detección robusta de personas. Los principales trabajos pueden dividirse en función de las técnicas utilizadas para extraer información de las imágenes. El primer grupo

incluye los detectores basados en segmentación [4] que hacen uso de un conocimiento a priori del fondo para separarlo del primer plano que correspondería a las personas a detectar. Este método presenta múltiples inconvenientes y es muy poco robusto a cambios de iluminación, fondos dinámicos y movimientos de la cámara. Por otro lado, los detectores basados en características de forma [5], [6] recurren al modelo de forma implícita o ISM (*Implicit Shape Model*) para obtener la localización de las personas en base a votaciones de un conjunto de sistemas de decisión basados en el aprendizaje de características similares a las de un diccionario previamente aprendido. Este método sólo funciona de manera adecuada con imágenes de alta resolución, obteniendo resultados muy pobres cuando la calidad de la imagen no es demasiado alta. Finalmente, los detectores basados en ventana deslizante proporcionan un mejor rendimiento para baja y media resolución. Esta técnica consiste básicamente en recorrer una imagen seleccionando cada vez una determinada región local, a la que se denominará ventana, y para la cual se extraerá un descriptor de características que permitirá saber si dicha región delimita parcial o totalmente el cuerpo de una persona o no. Para conocer si una ventana contiene una persona, se recurrirá al uso de un clasificador previamente entrenado [7],[8].

Dentro del planteamiento de detección basada en ventana deslizante, la selección de unos descriptores de características adecuados es fundamental para obtener unos resultados aceptables. A modo de clasificación del estado del arte, las características utilizadas en la detección de personas se pueden organizar en función de la información que proporcionan. Así, las clases básicas de las características son: las basadas en histogramas de orientación de gradientes (HOG) [9], [10], las características de forma [11], [12], las de movimiento [13], [14], las compuestas y, las basadas en articulaciones y partes. De entre todas ellas, HOG [9] ha resultado ser la que a nivel más básico ha proporcionado mejores resultados y sirve actualmente como referente en los sistemas de detección.

A pesar de las mejoras introducidas en los detectores, en el caso de los vídeos, la solución *frame a frame* no produce resultados completamente satisfactorios. Debido a ello, se hace necesario el uso de sistemas de seguimiento que mejoren las detecciones obtenidas y a su vez proporcionen un mecanismo de asociación con objetivos detectados previamente. Este problema se aborda con los llamados sistemas de seguimiento

mediante detección (*tracking by detection, TBD*). Para llevar a cabo esta tarea, es necesario recurrir a algún modelo de observación que permita combinar las informaciones de detección y seguimiento específicas de cada objetivo. Existen diferentes métodos probabilísticos de estimación (estimadores) que dan solución al problema de seguimiento de un objeto. Algunos de los más utilizados son: mínimos cuadrados recursivos (LMS), filtro de Kalman (KF) o de partículas (PF).

Finalmente, para la validación y comparación de los resultados de los diferentes trabajos, es habitual trabajar con bases de datos de imágenes difundidas entre la comunidad científica, que permitan hacer comparaciones entre distintas propuestas. Algunas de las más extendidas son KTH [16], Weizmann, INRIA [17], PETS, MuHAVI o CAVIAR [15], que es la usada en la propuesta presentada en esta comunicación.

Sin embargo, las imágenes incluidas en la mayoría de bases de datos no corresponden a situaciones habituales en escenarios de video-vigilancia, por lo que la solución aquí descrita ha sido también validada mediante una base de datos diseñada, grabada y anotada en situaciones realistas de video-vigilancia.

En este contexto, el trabajo presentado implementa una solución para la detección y seguimiento de múltiples personas en secuencias de video RGB, en aplicaciones video-vigilancia.

A continuación, en los apartados II y III se describe el sistema desarrollado para la detección de personas y su seguimiento, con asociación probabilística. Posteriormente, en el apartado IV se presenta la base de datos desarrollada para la validación del sistema propuesto. Finalmente, en el apartado V y VI se recogen los resultados experimentales obtenidos, así como las conclusiones y los trabajos futuros.

## II. DETECTOR DE PERSONAS

El detector de personas es el sistema que se encarga de determinar si en una imagen existen personas. En caso afirmativo, este deberá proporcionar tanto la posición como el tamaño en la imagen del individuo localizado. La estructura que tiene este bloque se muestra en la Fig. 1 y se compone de dos módulos: el de extracción de información y el clasificador.

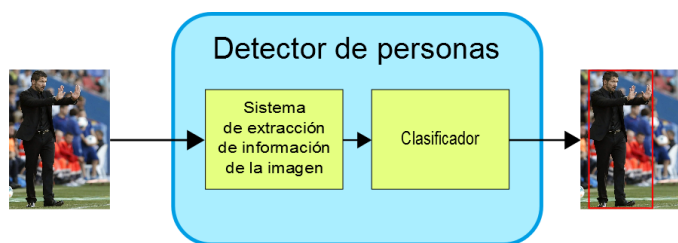


Fig. 1. Etapas del detector de personas.

Para la extracción de información de la imagen se han empleado descriptores HOG [9] (Histogramas de Gradientes Orientados) y un detector basado en ventana deslizante, que se ha implementado para una ventana de 64x128 píxeles. Además, para independizar el resultado de la escala de la persona se hace la extracción de descriptores realizando diferentes escalados de la imagen, en concreto analizando la imagen de entrada a 64 escalas diferentes.

Los descriptores obtenidos para las diferentes escalas se clasifican utilizando una SVM (Máquina de Soporte Vectorial), entrenada para la detección de personas con las imágenes de la base de datos de INRIA [17], que incluye un conjunto de imágenes de 64x128 píxeles de personas adecuadamente centradas y normalizadas.

El clasificador proporciona a su salida tanto la posición en la imagen en que se ha detectado cada persona (información del *bounding box* o caja de detección mediante las coordenadas de la esquina superior izquierda, su anchura y su altura, Fig. 2), como un valor de distancia al hiperplano de clasificación de la SVM, que está, por tanto, relacionado con la fiabilidad de la detección. Esta información de fiabilidad es de gran utilidad en la etapa posterior de seguimiento.



Fig. 2. Especificaciones del *bounding box*.

Al hacerse un análisis de la imagen a diferentes escalas, pueden generarse detecciones correspondientes a diferentes personas en las distintas escalas, o bien, a la misma persona en varias escalas. Cuando ocurre esto último, es necesario fusionar todas las detecciones correspondientes a la misma persona para presentarlas como una sola en la imagen original. Esto significa que, las detecciones únicas tienen que ser escaladas al tamaño que tiene la persona en la imagen real y las detecciones en distintas escalas, necesitan ser escaladas y después asociadas de modo que se muestren como una única detección. Para ello se aplica un algoritmo de particionamiento mediante el cual, se reconocen en la imagen todos los *bounding box* que hacen referencia a la misma persona y se fusionan quedando como resultado el más representativo, es decir, aquel que tiene la mayor coincidencia o peso de detección.

Las detecciones que arroja el sistema así diseñado constituyen las medidas del sistema completo, a incorporar como entrada de su etapa siguiente: la de seguimiento.

## III. SEGUIMIENTO CON ASOCIACIÓN PROBABILÍSTICA

El sistema de seguimiento implementado en este trabajo tiene un doble objetivo ya que: permite incrementar la robustez del sistema de detección al incorporar en él la información temporal en una secuencia de imágenes, y permite obtener información de la trayectoria en la escena seguida por la persona a lo largo de la secuencia.

Para el seguimiento de múltiples personas en imágenes se ha decidido emplear el filtro de Kalman, ya que proporciona buenos resultados con un bajo coste computacional. El filtro de Kalman [18] es un algoritmo iterativo de ejecución secuencial en dos etapas (predicción y corrección) que permite estimar y/o filtrar el estado de un sistema, en función del modelo establecido para el sistema las medidas reales que se obtengan en cada iteración.

Para modelar el movimiento de las personas se emplea un modelo de velocidad constante en el que se asume que la variación de la velocidad con que se mueven las personas en la escena es reducida. De esta forma, el modelo del sistema queda definido por la ecuación (1):

$$\mathbf{x}_{n+1} = \begin{bmatrix} x_{n+1} \\ y_{n+1} \\ v_{x_{n+1}} \\ v_{y_{n+1}} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_n \\ y_n \\ v_{x_n} \\ v_{y_n} \end{bmatrix} + \mathbf{w}_n \quad (1)$$

Es importante destacar que en general en la escena se encontrarán múltiples personas a seguir. Existen métodos que trabajan con un sólo estimador para predecir la posición de todos los objetos. No obstante, el filtro de Kalman en su forma más básica, no es capaz de conseguirlo. Esto obliga que cada persona a seguir en la escena necesite un filtro, que represente su hipótesis de seguimiento, conformándose así la etapa de seguimiento con un banco de filtros de Kalman. Para poder establecer la asociación entre los diferentes filtros del banco y las medidas arrojadas por la etapa de detección es necesario tener en cuenta las diferentes situaciones que pueden darse, analizadas a continuación.

#### A. Creación de un nuevo filtro de Kalman

Cuando aparece una nueva medida de detección (considerada como tal cuando no hay ninguna hipótesis con  $\mathbf{x}_n$  semejante a nivel de posición), el sistema debe elegir el momento adecuado para iniciar el seguimiento, es decir, de crear un nuevo filtro de Kalman. En este caso se ha decidido retrasar el inicio de iteración del filtro hasta que se hayan podido relacionar varias detecciones consecutivas con la hipótesis creada al efecto. Esta alternativa tiene la ventaja de que evita iniciar la iteración de filtros asociados a detecciones espurias, pero, retrasa el inicio del seguimiento de las detecciones correctas.

#### B. Asociación de datos

El proceso de asociación es una de las tareas fundamentales del bloque seguidor. En este trabajo se implementa un algoritmo de asociación basado en el vecino más próximo (*Nearest Neighbor Association, NN*), que calcula la distancia de posición en la imagen entre las medidas de salida del detector y la parte correspondiente de  $\hat{\mathbf{x}}_n$  de cada hipótesis para validar la asociación.

De este modo, el algoritmo de asociación, para cada medida, calcula la distancia euclídea con todas las hipótesis posibles, tras lo cual compara las distancias con un umbral de cercanía. En este punto, toda asociación que cumpla ese umbral es considerada candidata como válida. Finalmente, el algoritmo NN validará la asociación de menor distancia.

Al realizar el algoritmo para todas las medidas recibidas de la etapa de detección, el sistema consigue diferenciar aquellas que proceden de personas que ya están siendo seguidas por un filtro de Kalman (hipótesis ya validadas) de las detecciones que no tienen hipótesis de seguimiento (KF) aún no iniciado.

#### C. Iteración de seguidores

Una vez que el algoritmo de asociación ha finalizado, cada filtro puede estar en dos situaciones en función de si ha tenido, o no, nuevas medidas asociadas en su iteración anterior.

El primer caso es el común cuando la persona está siendo correctamente seguida en una secuencia de imágenes. En esta situación, el filtro corrige su estimación en cada iteración, tras lo cual realiza una nueva predicción del estado  $\mathbf{x}_n$ .

Si en una iteración el filtro no tiene nuevas medidas asociadas a su estimación de posición, es posible que sea debido a que la persona cuya detección seguía se haya ido de la escena o haya desaparecido tras algún objeto (oclusión). En estas situaciones, el estimador no dispone de datos para ejecutar la etapa de corrección, por tanto, únicamente ejecutará la de predicción. Cuando la persona vuelva a ser encontrada por el detector, se reanudará el proceso iterativo del sistema de seguimiento relacionado (corrección y predicción).

#### D. Eliminación de filtros

Cuando una persona desaparece de la escena y no se vuelve a detectar (no existen nuevas medidas), se ha de finalizar el estimador relacionado a dicha persona.

Para ello se procede del mismo modo que en el proceso de creación: se mantiene la etapa de predicción durante algunas iteraciones, eliminando la hipótesis definitivamente si pasadas éstas no se le vuelve asociar ninguna medida, implicando que la persona ya no está en la escena.

## IV. BASE DE DATOS GBA2016 PARA APLICACIONES DE VIDEO-VIGILANCIA EN ESCENARIOS REALISTAS

Como ya se ha comentado en la introducción, a pesar de que existen numerosas bases de datos disponibles para la validación de las propuestas de detección de personas, ninguna cumple los requisitos necesarios por este trabajo. Debido a ello, se ha creado y etiquetado una nueva base de datos, que incluye situaciones realistas para aplicaciones de video-vigilancia, con objeto además de ponerla a disposición de la comunidad científica para su utilización y expansión. Las principales características de la base de datos se presentan a continuación.

La base de datos GBA2016 consta de unas 280 secuencias de imágenes diferentes, adquiridas en un entorno interior, con una cámara comercial GoPro Hero 4, con una resolución de 1280x720 píxels a 50 fps.

En las secuencias aparecen diferentes personas realizando acciones cotidianas que pueden encontrarse en aplicaciones de vigilancia (como andar, correr, sentarse o caerse), que se plantea ampliar con más grabaciones con nuevas personas y acciones de interés (como puede ser pelearse, saltar, etc.)

En las grabaciones, la cámara utilizada se ubica en una posición elevada para reducir, en la medida de lo posible, las oclusiones entre diferentes personas.

La Fig. 3 muestra el plano y una fotografía de la zona en que se han realizado las grabaciones, en las que se ha marcado la región de interés considerada.

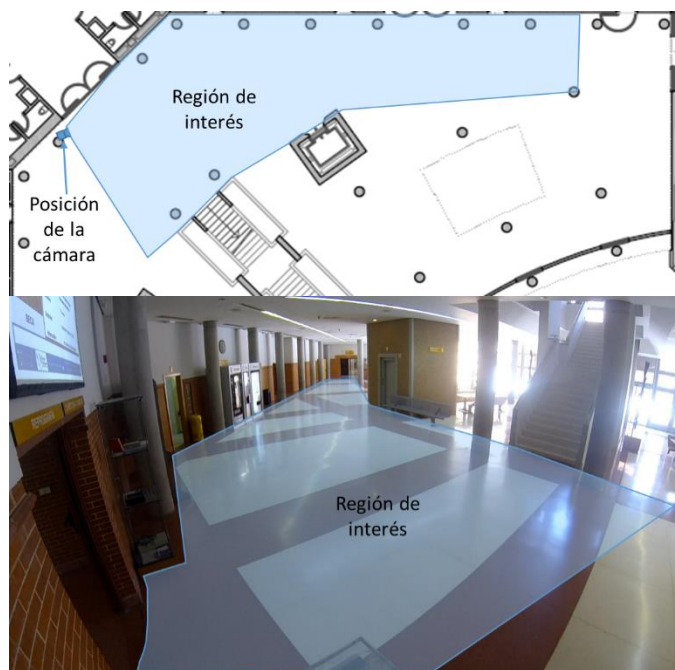


Fig. 3. Zona de interés en la que se han realizado las grabaciones de la base de datos GBA2016.

Cabe destacar que se han llevado a cabo grabaciones en días y horarios diferentes, para tratar de incluir cambios ambientales debidos, sobre todo, a las diferencias en la iluminación exterior.

Además, se han grabado dos tipos de secuencias diferentes. El primer tipo incluye aquellas secuencias en las que aparece una sola persona en cada imagen, mientras que en el segundo tipo se agrupan las secuencias que incluyen diferentes personas en una imagen llevando a cabo acciones distintas o iguales.

## V. RESULTADOS EXPERIMENTALES

A continuación se presentan los resultados experimentales obtenidos para el sistema desarrollado. Para obtener estos resultados se ha realizado la implementación en C++ de los algoritmos de detección y seguimiento descritos, empleando las librerías OpenCV [19] en un entorno basado en Linux. Se han realizado pruebas experimentales tanto con imágenes pertenecientes a la base de datos de CAVIAR [15], ampliamente utilizada en la literatura (formado por secuencias grabadas en un centro comercial de Portugal con una resolución de 384x288 píxeles a 25 fps), como con imágenes pertenecientes a la base de datos grabada y etiquetada para esta aplicación GBA2016. En la Tabla I se muestra el número de secuencias, así como el número medio de imágenes por cada secuencia para ambas bases de datos.

En los siguientes apartados se muestra el funcionamiento de cada bloque del sistema expuesto por separado. De este modo se puede analizar el efecto que tienen en el sistema completo los distintos bloques que lo forman (detección y seguimiento).

TABLA I  
NÚMERO DE SECUENCIAS E IMÁGENES INCLUIDAS EN LAS BASES DE DATOS UTILIZADAS.

	CAVIAR	GBA16s
<i>Secuencias utilizadas</i>	26	15
<i>Número medio de imágenes por secuencia</i>	1407	5582
<i>Número total de imágenes</i>	36582	83730

Para considerar el análisis de tiempos de ejecución hay que tener en cuenta que las pruebas experimentales se han realizado en un ordenador con procesador Intel Core 2 y 2GB de RAM

### A. Resultados del detector de personas

En la Fig. 4 se muestran los resultados de la detección de personas en dos imágenes pertenecientes a la base de datos de CAVIAR. En ambas imágenes se pueden observar las cajas alrededor de las personas detectadas. El color de las diferentes cajas está relacionado con la fiabilidad de la detección arrojada por el clasificador SVM. Los colores utilizados, ordenados de menor a mayor fiabilidad, son los siguientes: rojo, naranja, amarillo, verde azul y magenta.

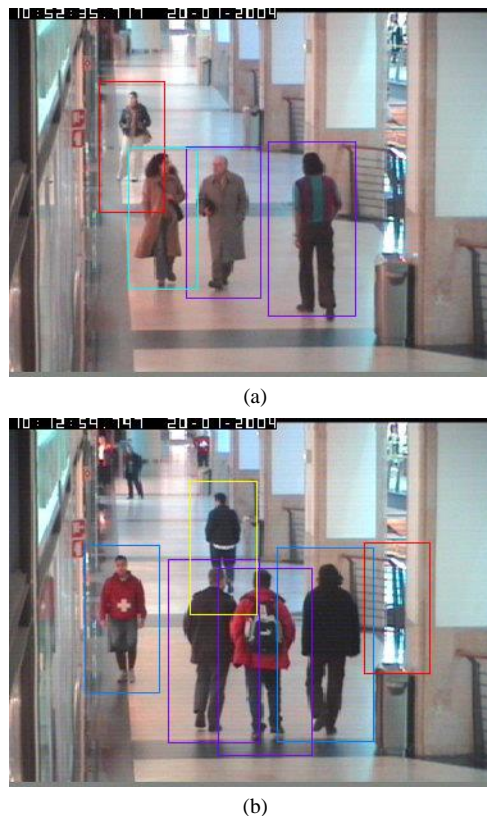


Fig. 4. Resultados de detección obtenidos para imágenes pertenecientes a la base de datos de CAVIAR.

En la Fig. 5 se presenta el resultado de la detección para una imagen perteneciente a la base de datos GBA2016. Dado que estas imágenes tienen una mayor resolución, se detectan personas a una distancia de la cámara mayor que usando las imágenes de CAVIAR.

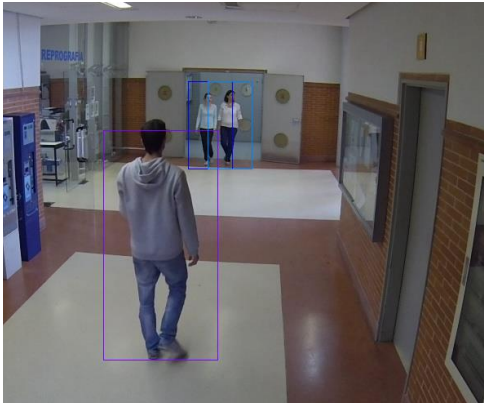


Fig. 5. Resultado de la detección obtenidos para una imagen perteneciente a la base de datos GBA2016.

Se ha analizado también el tiempo de ejecución medio consumido por la etapa de detección para ambos ejemplos, siendo los valores obtenidos con el hardware de referencia de 183.66 ms para las imágenes de CAVIAR y 401.28 ms para las de GBA2016. Los valores del algoritmo completo se analizan en la sección C de este apartado.

### B. Resultados del sistema de seguimiento

La incorporación del banco de filtros de Kalman permite aumentar la robustez del sistema, especialmente en los casos en que hay errores en la detección o si aparecen oclusiones, como demuestran las siguientes pruebas.

En la Fig. 6 se muestra el resultado de incluir la etapa de seguimiento cuando existe una oclusión parcial debido a que dos personas se cruzan en la escena. Se puede observar cómo en el caso de la persona ocluida no se obtiene resultados del detector (cruces amarillas) pero si de la etapa de predicción del filtro de Kalman (cruces azules).



Fig. 6. Efecto de la etapa de seguimiento ante oclusiones.

Por otro lado, en la Fig. 7 se presenta el efecto del seguidor en un caso en que temporalmente se produzca un error de detección (falso negativo, caja amarilla). Mientras esto ocurre, el seguidor mantiene la predicción de la hipótesis (cruces amarillas) hasta que finalmente vuelve a haber detección (cruces azules).

En relación al tiempo de cómputo del algoritmo de seguimiento, se concluye que éste no depende de la resolución de las imágenes, como cabía esperar, siendo su valor medio de 0.723ms.



Fig. 7. Efecto del filtro de Kalman ante la pérdida temporal de la detección.

### C. Tiempo de procesamiento de los algoritmos

Finalmente, se ha analizado el tiempo de ejecución consumido por el algoritmo global, para las dos bases de datos consideradas. En la Tabla I se resumen los resultados obtenidos. Los valores expuestos son los valores mínimos, medios y máximos del tiempo consumido por las etapas de detección y seguimiento al procesar las secuencias indicadas en la Tabla II.

TABLA II  
TIEMPOS DE PROCESAMIENTO CONSUMIDO POR EL ALGORITMO (EN MILISEGUNDOS) Y NÚMERO DE IMÁGENES PROCESADAS POR SEGUNDO.

	CAVIAR	GBA16s
<i>Tiempo mínimo (ms)</i>	173.02 ms	380.83 ms
<i>Tiempo medio (ms)</i>	184.44 ms	402.01 ms
<i>Tiempo máximo (ms)</i>	266.52 ms	565.86 ms
<i>Número medio de imágenes por segundo (fps)</i>	5.42 fps	2.49 fps

Los valores mostrados en la Tabla II no permiten la ejecución en tiempo real del algoritmo de detección, sin embargo, estos tiempos pueden reducirse de forma significativa mejorando las prestaciones del sistema de procesamiento utilizado y realizando la ejecución del detector a diferentes escalas en paralelo empleando la tarjeta gráfica (GPU, Graphics Processing Unit), que cuenta con la arquitectura de cálculo paralelo CUDA (Compute Unified Device Architecture) y las herramientas asociadas [20], en combinación con el procesador y las funciones actuales.

## VI. CONCLUSIONES Y TRABAJOS FUTUROS

En este trabajo se ha presentado una solución robusta para la detección y seguimiento de un número variable de personas, a partir de secuencias de imágenes RGB, en un entorno realista, orientado a aplicaciones de video-vigilancia. El sistema desarrollado permite tanto la detección de personas en imágenes, como el conteo de las mismas lo que facilita su uso en diferentes aplicaciones como la video-vigilancia o el control de aforos.

Para la detección de personas se ha empleado un descriptor HOG a 64 escalas diferentes y un clasificador SVM lineal. Además, se ha incorporado a la salida de detección un seguidor para múltiples personas basado en un banco de filtros de Kalman que permiten incrementar la robustez del sistema frente a oclusiones o pérdidas temporales de la detección.

Para la validación del algoritmo desarrollado se han realizado múltiples pruebas experimentales empleando tanto la base de datos de imágenes del proyecto CAVIAR, ampliamente utilizada por la comunidad científica, como una base de datos más realista diseñada, grabada y etiquetada para este tipo de aplicaciones.

Los resultados obtenidos han permitido validar el proceso de detección, así como demostrar que la incorporación de la etapa de seguimiento incrementa la robustez frente a oclusiones o errores en la detección, con un aumento mínimo de la carga computacional.

Se ha analizado también el tiempo de cómputo del algoritmo, concluyendo que con el hardware disponible no es posible ejecutarlo en tiempo real, y arrojando como trabajo futuro la implementación del detector en GPU para aprovechar su capacidad para el cálculo en paralelo de los descriptores y su posterior clasificación SVM. Con esta implementación, así como con una mejora de las características del procesador utilizado se espera reducir de forma significativa el tiempo de proceso de modo que sea posible su ejecución en tiempo real.

Por otro lado, se espera poner a disposición de la comunidad científica la base de datos desarrollada (imágenes y etiquetado) en los próximos meses.

## VII. AGRADECIMIENTOS

Este trabajo ha sido financiado por el Ministerio de Economía y Competitividad a través del proyecto SPACES-UAH (TIN2013-47630-C2-1-R), y por la Universidad de Alcalá a través de los proyectos DETECTOR (CCG2015/EXP-019) y ARMIS (CCG2015/EXP-054).

## REFERENCIAS

- [1] R. Pope, "A survey on vision-based human action recognition", in *Image and Vision Computing*, 2010.
- [2] R. Arroyo, J. J. Yebes, L. M. Bergasa, I. G. Daza, and J. Almazán, "Expert video-surveillance system for real-time detection of suspicious behaviors in shopping malls," *Expert Systems with Applications*, vol. 42, no. 21, pp. 7991 – 8005, 2015.
- [3] Carlos Martínez, Marcos Baptista, Cristina Losada, Marta Marrón y Valeria Boggian. "Human Action Recognition in Realistic Scenes Based on Action Bank". *Proceedings of the International Work-Conference on Bioinformatics and Biomedical Engineering (IWBBIO 2016)*, pp: 314-325. 2016.
- [4] C. Gu, J. Lim, P. Arbelaez, and J. Malik, "Recognition using regions," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, vol. 2, Junio 2009, pp. 1030–1037.
- [5] B. Leibe, E. Seemann, and B. Schiele, "Pedestrian detection in crowded scenes," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1, Junio 2005, pp. 878–885.
- [6] K. M. Edgar Seemann, Bastian Leibe and B. Bernt Schiele, "An evaluation of local shape-based features for pedestrian detection," in *British Machine Vision Conference*, Septiembre 2005.
- [7] C. Papageorgiou and T. Poggio, "A trainable system for object detection," *International Journal of Computer Vision*, vol. 38, no. 1, pp. 15–33, Junio 2000.
- [8] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, Mayo 2004.
- [9] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1, Junio 2005, pp. 886–893.
- [10] Q. Zhu, M.-C. Yeh, K.-T. Cheng, and S. Avidan, "Fast human detection using a cascade of histograms of oriented gradients," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 2, 2006, pp. 1491–1498.
- [11] D. Gavrila and V. Philomin, "Real-time object detection for smart vehicles," in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, vol. 1, 1999, pp. 87–93.
- [12] D. Gavrila, "A bayesian, exemplar-based approach to hierarchical shape matching," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 8, pp. 1408–1421, Aug 2007.
- [13] P. Viola, M. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, Octubre 2003, pp. 734–741.
- [14] N. Dalal, B. Triggs, and C. Schmid, "Human detection using oriented histograms of flow and appearance," in *Proceedings of the 9th European Conference on Computer Vision*, vol. 2, 2006, pp. 428–441.
- [15] CAVIAR Dataset <http://homepages.inf.ed.ac.uk/rbf/CAVIAR/> (último acceso febrero 2016)
- [16] Christian Schuldt, Ivan Laptev and Barbara Caputo. *Recognizing Human Actions: A Local SVM Approach*, Proc. ICPR'04, Cambridge, UK. 2004.
- [17] "Página de la base de datos de INRIA," <http://pascal.inrialpes.fr/data/human/>. (último acceso febrero 2016)
- [18] R.E. Kalman. "A new approach to linear filtering and prediction problems", *Transactions of the ASME, Journal of basic engineering*, 82:35-45. 1960.
- [19] Bradski, Gary and Kaehler, Adrian. *Learning OpenCV: Computer vision with the OpenCV library*. O'Reilly Media, Inc. 2008.
- [20] Página web acerca de CUDA: <https://developer.nvidia.com/about-cuda> (último acceso mayo 2016)