

Detección de objetos en secuencias de video mediante modelos estadísticos y neuronales

Miguel A. Molina-Cabello¹, Ezequiel López-Rubio¹ y Rafael Marcos Luque-Baena²

¹ Departamento de Lenguajes y Ciencias de la Computación. Universidad de Málaga. Bulevar Louis Pasteur, 35. 29071 Málaga. España.

² Departamento de Ingeniería de Sistemas Informáticos y Telemáticos. Universidad de Extremadura. Centro Universitario de Mérida. 06800 Mérida. España. {miguelangel,ezeqlr}@1cc.uma.es, rmluque@unex.es

Resumen El uso de un mayor número de videos e imágenes, incluyendo los obtenidos por cámaras con movimiento (PTZ), hace necesario contar con aplicaciones que procesen automáticamente esta información. Sin embargo, uno de los principales problemas para elaborar estas aplicaciones es detectar los objetos que pertenecen al primer plano, máxime si la cámara es PTZ. En este trabajo se propone el desarrollo de una tesis doctoral que profundice en esta temática, definiendo sus objetivos, la metodología aplicada y el plan de trabajo establecido.

Keywords: Visión por computador, procesamiento de imágenes, modelado del fondo, detección de objetos, cámaras PTZ

1. Introducción

El número de videos e imágenes existentes es cada vez mayor debido al aumento del uso de cámaras de video. Normalmente son cámaras 2D, aunque existen otros tipos como 3D, multispectrales, con movimiento (también llamadas Pan-Tilt-Zoom o PTZ), etc.

Debido a dicha cantidad de datos (*Visual big data*) se hace necesario realizar un procesamiento automático de imágenes ya que gestionarlos de manera rigurosa manualmente sería muy costoso, siendo deseable contar con aplicaciones rápidas (máxime si dicho procesamiento es en tiempo real) que puedan reconocer el comportamiento de los actores que participan en la escena capturada en una imagen y desarrollar otras tareas a partir de este reconocimiento. Así, centrándonos en la videovigilancia, un sistema automático podría identificar personas, hacerles un seguimiento, entender su comportamiento, detectar objetos perdidos, actividades sospechosas o situaciones de peligro, activar alarmas... Incluso podría practicar el seguimiento de los objetos o personas realizando los movimientos de cámara necesarios si ésta fuera del tipo PTZ.

Además, sería interesante contar con un buscador inteligente que permitiera la recuperación en bases de datos de imágenes y videos basada en su contenido (*Content based retrieval*), pudiendo satisfacer búsquedas como, por ejemplo: "encontrar todos los videos en los que aparezca un gato".

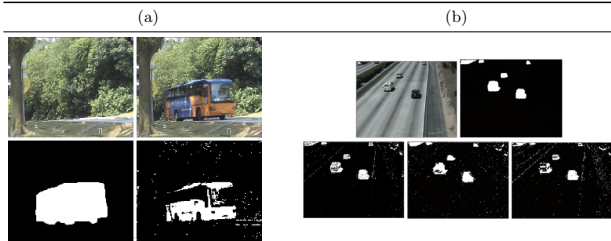


Figura 1. Ejemplos de segmentación. (a) Arriba: escena e imagen con un objeto presente en dicha escena; abajo: segmentación ideal y segmentación obtenida con un algoritmo de detección de objetos. (b) Arriba: imagen a segmentar y su segmentación ideal; abajo: resultado producido por tres algoritmos distintos.

2. Definición del problema

Este procesamiento de teledetección se puede descomponer en tres fases: una primera etapa de detección de los objetos en movimiento (en primer plano), una segunda fase que controla el seguimiento de dichos objetos y una tercera y última acción que consiste en el análisis del comportamiento.

Sin embargo, todo este procesamiento automático es el mayor inconveniente que existe actualmente, centrándonos principalmente en la primera fase, donde es crítico detectar el fondo y los objetos que aparecen en ellas, utilizando información temporal de la secuencia de vídeo capturada.

Existen diversos algoritmos para solventar el problema pero no existe ninguno que resuelva adecuadamente todas las situaciones posibles. Esto se debe a que no es trivial que un algoritmo pueda gestionar situaciones inesperadas que puedan aparecer y obtener unos resultados fiables. Hay multitud de dificultades que entorpecen nuestra labor, como objetos estacionarios en movimiento, ya sea por objetos de fondo móviles (por ejemplo, olas del mar) o por la oscilación de la cámara, camuflaje (los objetos de primer plano tienen colores muy similares al fondo de la escena), cambios de iluminación, aparición de sombras, ruido, etc.

Dos ejemplos de la aplicación de algoritmos de detección del fondo y objetos en primer plano pueden observarse en la Figura 1. En (a) existen zonas que pertenecen al objeto y que el algoritmo empleado no ha reconocido. Por otro lado, existe ruido como consecuencia de sombras y el cambio de posición de algunos elementos, como las ramas de los árboles, motivados supuestamente por el cambio de la posición del sol con respecto a la escena y por la acción del viento. Por su parte, en (b) se puede observar que el resultado producido por cada algoritmo es distinto, presentando distinta calidad.

Esta problemática es independiente del tipo de cámara utilizada, si bien es cierto que la dificultad aumenta si en lugar de una cámara común se utiliza una

del tipo PTZ debido a que su movimiento no hace posible que pueda visualizarse la escena completa en cualquier instante de tiempo.

3. Estado del arte

La visión por computador es una de las áreas más demandadas actualmente en el mundo de la investigación, existiendo multitud de revistas que solicitan publicaciones de esta índole [8].

Nuestro cometido se centra en el área del procesamiento de vídeo, principalmente en los métodos de detección de objetos y fondo de una imagen. Podemos citar algunas publicaciones de relevancia como, por ejemplo, el método [9] (hace uso de una distribución gaussiana) o el modelo [5] (redes neuronales). También destacamos la publicación [6] perteneciente al campo del seguimiento de objetos.

Nuestra investigación se basará en trabajos previos ya desarrollados por nuestro grupo de investigación de los que podemos citar el algoritmo de aproximación estocástica con uso de mixturas de distribuciones uniformes y gaussianas [2]; el modelo [3] basado en el uso de redes neuronales autoorganizadas y distribuciones gaussianas para detectar objetos en una imagen, concretamente imágenes de tráfico; o la propuesta [4] de un espacio de color y un proceso de selección del componente color ponderado para la detección de objetos y fondo en vídeos mediante el uso de mapas autoorganizados.

Para la realización de estudios sobre la bondad de los nuevos modelos desarrollados nos basaremos en la comparación con otros del estado del arte. Para este cometido se puede hacer uso de paquetes que incorporan modelos ya implementados, como es la *librería BGS* [7]. Además de las distintas medidas utilizadas en este tipo de estudios, necesitamos vídeos de los que tengamos su segmentación ideal (*Ground Truth Mask*), existiendo librerías de vídeos para este propósito, como el *2014 dataset* de ChangeDetection.net [1].

4. Objetivos

Nuestro principal objetivo es el de desarrollar (mediante el uso de técnicas como redes autoorganizadas, modelos estadísticos o la aproximación estocástica) nuevos modelos y algoritmos que permitan detectar con mayor calidad objetos de primer plano en las secuencias de imágenes analizadas, poniendo especial interés en los vídeos captados por cámaras PTZ ya que el proceso de modelado del fondo de la escena es más complejo y las investigaciones existentes son prácticamente nulas en comparación con las de cámaras tradicionales. Para ello se estudiarán diversas técnicas existentes y se propondrán mejoras. Una de estas técnicas que podemos citar es la generación del modelo del fondo mediante una imagen panorámica que se va elaborando conforme la cámara se va moviendo. Un ejemplo de dicha técnica puede verse en la Figura 2.

También nos hemos marcado como reto desarrollar algoritmos de seguimiento de objetos mediante cámaras PTZ, de manera que la cámara se mueva automáticamente en función del movimiento de los objetos presentes en la escena.



Figura 2. Ejemplo de generación de panorámica de un video de cámara PTZ. (a) y (b) se corresponden con los frames 6 y 339, respectivamente, mostrando la imagen captada por la cámara en ese instante y la panorámica elaborada hasta dicho momento.

5. Metodología

Los algoritmos de detección de objetos, además de ofrecer una solución al problema, deben tener una complejidad baja para ejecutarse en un tiempo razonable, máxime en un sistema en tiempo real. Así, la calidad del método está determinada por la fiabilidad de la solución y por el tiempo que consume.

La base de nuestra investigación es el método científico, donde todo proyecto que lo cumpla estará sujeto al principio de reproducibilidad (poder repetir el experimento) y refutabilidad (poder demostrar que la hipótesis es falsa). También se hará uso de una metodología iterativa e incremental, así como de una implementación adecuada para hacer que el proyecto sea extensible, modificable...

Por otro lado, se establecerán criterios de evaluación, utilizando medidas de rendimiento calculadas entre el resultado ideal y el resultado producido por nuestro algoritmo, tales como el acierto, la precisión, la F-medida... Además, se realizará una comparación con otros modelos mediante un estudio de las diferencias de los resultados producidos por nuestro método con respecto a otros del estado del arte, basándonos en las medidas de rendimiento establecidas anteriormente.

Concretando técnicamente la detección de objetos mediante cámaras PTZ, se generará un fondo panorámico utilizando técnicas de emparejamiento de puntos entre fotogramas consecutivos. Posteriormente se identificará la posición en el fondo panorámico a la que se corresponde el fotograma analizado. Por último, se aplicará el modelo de aprendizaje sobre esa región del panorama para detectar el movimiento y actualizarlo con la información obtenida.

6. Plan de trabajo

La presente tesis fue iniciada en octubre de 2014 y estaba planificada para 5 años puesto que en un principio el doctorado se iba a realizar con una dedicación a tiempo parcial, aunque finalmente se prevé que se realice en 4 años.

Podemos esbozar distintas actuaciones a realizar durante nuestro doctorado. Primero se abordará un aprendizaje de la herramienta de trabajo. Posteriormente se estudiará el estado del arte, analizando algoritmos y métodos de muestra temática, además del estudio de los modelos desarrollados por el grupo de investigación. A continuación se generarán y desarrollarán nuevos modelos para

finalmente proceder para cada uno de ellos a la evaluación de resultados y extracción de conclusiones, donde se elaborará una comparativa de los resultados obtenidos por los nuevos modelos con los de los algoritmos existentes, documentando qué características tienen nuestros nuevos modelos.

En un primer momento los modelos desarrollados serán de cámara estacionaria y después, una vez estudiada la temática en profundidad, de cámaras PTZ. En función del devenir de estos trabajos se continuará con investigaciones sobre el seguimiento de objetos con cámaras PTZ.

Agradecimientos

Los autores agradecen que esta tesis se encuentre financiada parcialmente por el Gobierno Autónomo de Andalucía (España) bajo el proyecto TIC-657 Sistemas autoorganizados y estimadores robustos para videovigilancia. También tiene colaboración del Ministerio de Economía y Competitividad de España: proyecto TIN2014-53465-R; del Gobierno Autónomo de Andalucía (España): proyecto TIC-6213; y del Gobierno Autónomo de Extremadura (España): proyecto IB13113. Todos estos proyectos incluyen fondos del Fondo Europeo de Desarrollo Regional (European Regional Development Fund, ERDF). Por último, agradecer los recursos y asistencia proporcionados por el SCBI (Supercomputing and Bioinnovation Center) de la Universidad de Málaga.

Referencias

1. ChangeDetection.net: 2014 dataset. <http://changedetection.net/>, [Online; accessed 12-Apr-2016]
2. López-Rubio, E., Luque-Baena, R.M.: Stochastic approximation for background modelling. *Computer Vision and Image Understanding* 115(6), 735–749 (2011)
3. Luque-Baena, R., López-Rubio, E., Domínguez, E., Palomo, E., Jerez, J.: A self-organizing map to improve vehicle detection in flow monitoring systems. *Soft Computing* (2015)
4. López-Rubio, F., Domínguez, E., Palomo, E., López-Rubio, E., Luque-Baena, R.: Selecting the color space for self-organizing map based foreground detection in video. *Neural Processing Letters* (2015)
5. Maddalena, L., Petrosino, A.: A self-organizing approach to background subtraction for visual surveillance applications. *IEEE Transactions on Image Processing* 17(7), 1168–1177 (2008)
6. Smeulders, A., Chu, D., Cucchiara, R., Calderara, S., Dehghan, A., Shah, M.: Visual tracking: An experimental survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36(7), 1442–1468 (2014)
7. Sobral, A.: Bgs library. <https://github.com/andrewwsobral/bgslibrary/>, [Online; accessed 12-Apr-2016]
8. University, H.O.: Journal special issues. <http://dsmc2.eap.gr/publications/journals/si#PERCEPTVP>, [Online; accessed 25-Apr-2016]
9. Wren, C.R., Azarbayejani, A., Darrell, T., Pentland, A.P.: Pfindex: Real-time tracking of the human body. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 19(7), 780–785 (1997)