UABDIVULGA
BARCELONA RECERCA I INNOVACIÓ

01/2010

Video-Hermenéutica o la interpretación del comportamiento humano en secuencias de imágenes

La hermenéutica, definida como el arte de interpretar un mensaje, se centró durante muchos siglos en teorizar el proceso de interpretación de textos escritos, sobre todo bíblicos. La aparición, a finales del s. XIX, de las primeras grabaciones de secuencias de imágenes amplió el dominio de búsqueda, interesando durante el s.XX a filósofos como Heidegger: en este nuevo formato de comunicación humana (visual), el texto cinematográfico se convertía en un juego interpretativo donde el lenguaje visual se articulaba en una red de múltiples lecturas. La hermenéutica de secuencias de imágenes, o Vídeo-Hermenéutica, implica explicar el valor subjetivo y social del comportamiento humano observado en secuencias de imágenes y, en general, de todo contenido multimedia. La vídeo-hermenéutica no sólo se interesa por lo que pasa en un vídeo, sino por entender qué significado tiene lo que está siendo descrito, qué mensaje nos transmite como observadores.

Este análisis del comportamiento humano en secuencias de imágenes también se ha modelizado en términos computacionales dentro del campo de las Ciencias de la Computación. Y esto ha sido posible gracias a los adelantos técnicos y de hardware: en particular, al abaratamiento de los costes de las cámaras, que comportó una expansión de su uso para la video-vigilancia. Y este uso ha provocado la necesidad de analizar automáticamente y en tiempo real el comportamiento humano observado en millones de cámaras.

Estos factores han generado importantes aportaciones científico-técnicas también en las áreas de la Visión por Computador y de la Inteligencia Artificial: un compendio de los trabajos más recientes en este ámbito se encuentran en un número especial del *International Journal of Pattern Recognition and Artificial Intelligence*, el cual es introducido en el artículo descrito en las referencias. A grandes rasgos, se pueden categorizar 4 grados de complejidad en los sistemas de visión artificial, como los hitos que se han logrado en las dos últimas décadas. El primer paso, básico y crítico a la vez, corresponde a la detección y seguimiento del (i) *movimiento*. Sin detección no hay interpretación, y surgen muchos problemas en este sentido: saturaciones, sombras, camuflaje, movimiento del fondo, ocultaciones... La siguiente tarea corresponde al reconocimiento de las (ii) *acciones* realizadas por los agentes detectados, por ejemplo andar, agacharse o correr. El tercer grado implica modelitzar las (iii) *actividades*, definidas como acciones más interacciones y reacciones. Las actividades determinan si, por ejemplo, dos agentes se acercan, giran, se persiguen.... Por último, la categoría con más carga contextual corresponde a los (iv) *comportamientos*, que sitúan las actividades en una escena concreta; por ejemplo, si dos agentes se acercan rápidamente entre ellos (actividad), la escena es una calle y un agente es detectado como humano y el otro como vehículo, se interpretará un posible peligro de atropello.

Esta gradación nos indica cómo se va profundizando en la semántica de cada imagen para reducir la incertidumbre y el error del proceso interpretativo y mejorar así la utilidad semántica de la explicación del comportamiento observado. Es así como se va dando cada vez más importancia a la identificación, de una parte, de la escena donde se produce el movimiento y de sus regiones semánticamente más relevantes, como por ejemplo una acera, una parada de autobús o un paso de peatones, y por otra parte de aquellos objetos con los que los agentes detectados pueden interactuar, como por ejemplo bolsos, bicicletas, sillas, puertas, ventanas... Por último, se están diseñando estrategias que incorporan otros elementos semánticos que pueden aparecer en el vídeo y que son adicionales a la información que hay a la imagen, como por ejemplo el audio o la aparición de palabras escritas.

Las tendencias más prometedoras en este sentido indican que el futuro de la vídeo-hermèutica se encuentra, entre otros, en Internet. Poruna parte, debido a la enorme variedad de comportamientos humanos y de escenas, muchísimo más alta que en la vídeo-vigilancia, una estrategia emergente consiste en utilizar las imágenes y vídeos que los usuarios suben a la red para modelizar esta variedad. Por otra parte, se están diseñando sistemas que mejoran el análisis de, no sólo el contenido multimedia generado por empresas audiovisuales y otras industrias culturales, sino del contenido *streaming* generado por plataformas como *YouTube*.

Ha pasado más de un siglo desde que pioneros como Muybridge y Marey hicieron las primeras grabaciones de movimientos humanos con finalidades de análisis; en aquella época una de las aplicaciones más demandadas era poder determinar la distribución más óptima del material

bélico que traía un soldado, para reducirle la fatiga durante jornadas muy largas. En nuestros días, la aplicación más prometedora a largo plazo es el diseño de programas de anotación automática de vídeo para aplicar indexaciones más informativas y eficientes de archivos multimedia, para refinar los resultados de los motores de busca y, en definitiva, para encontrar rápidamente el contenido semántico deseado en un volumen de datos cada vez más infinito.

Jordi González
poal@cvc.uab.es

Referencias

"Video Analysis and Understanding for Surveillance Appilcations". Wang, Liang; Wu, Qiang; Li, Ming; Gonzalez, Jordi; Geng, Xin. International Journal of Pattern Recognition and Artificial Intelligence, 23 (7): 1221-1222 NOV 2009.

View low-bandwidth version