

OpenSourceVS: BIBLIOTECA DE CÓDIGO ABIERTO PARA VIDEO SINOPSIS

José Olivera, Eduardo Oliva, Nelson Ponzoni Cuadra

Instituto de Investigación en Señales, Sistemas e Inteligencia Computacional, sinc(i) - CONICET / Facultad de Ingeniería y Ciencias Hídricas

Área temática: Ingenierías.

Sub-área: Informática y Sistemas de Información.

INTRODUCCIÓN

Los sistemas de video-vigilancia han tenido, en los últimos años, un crecimiento continuo en popularidad y disponibilidad, que lleva a que se capturen videos las 24hs en infinidad de cámaras alrededor del mundo. Si bien esto permite tener un registro completo de los eventos sucedidos, se hace cada vez más importante el problema de la gestión eficaz del almacenamiento y el análisis de la información almacenada (Thounaojam, 2014). La video sinopsis consiste en generar videos de corta duración que contengan únicamente la información relevante proveniente de videos extensos, a partir de ubicar los objetos de interés y fundir su aparición y movimiento en un mismo espacio-tiempo (Fu et al., 2014).

Los métodos del estado del arte han ido evolucionando desde métodos simples como eliminar cuadros (en inglés, frames) cada cierto tiempo sin fijarse en la actividad de la escena (Nam & Tewfik, 1999), y métodos adaptativos que eliminan cuadros de manera selectiva en períodos de inactividad (Petrovic et al., 2005). Los métodos más recientes están basados en minimizar una función de energía, en cuya función de costo se ponen en juego restricciones de actividad, duración de la sinopsis, fragmentación de objetos, entre otros (Rav-Acha et al., 2006; Yao et al., 2014).

Si bien en la literatura se reportan estas técnicas, no se ha presentado y puesto a disposición –hasta nuestro conocimiento– una herramienta software que implemente estos métodos y pueda ser utilizada no sólo para la comunidad científica, sino también por la comunidad en general interesada en esta tecnología y sus aplicaciones. En este trabajo se presenta una biblioteca que implementa un sistema de video sinopsis basada en la detección y segmentación de objetos móviles, su extracción y fusión para generación del video-resumen sobre el fondo modelado.

VIDEO SINOPSIS BASADA EN OBJETOS

A continuación se presentan los detalles del método y de la biblioteca desarrollada que permite obtener un resumen de video aplicando técnicas de procesamiento digital de imágenes.

Modelado del fondo

El objetivo de esta etapa es obtener el fondo estático de la escena, en una imagen que se denomina de “sólo fondo”. Esta imagen será utilizada para extraer los objetos “móviles” de la escena y posteriormente, servirá como fondo para superponer los objetos detectados que se deseen, de forma solapada en el tiempo o no. Este fondo se modela, para cada pixel p , mediante una mezcla de Gaussianas $\bar{\eta}_p = (\bar{N}_p, \bar{\sigma}_p^2)$. A los fines prácticos de este

Proyecto: CAI+D 2001 #58-511, código PI 501201100511LI

Director del proyecto: César Martínez

Directores de los autores: César Martínez y Enrique Albornoz

trabajo, se define $\bar{\eta}_p = (\mathcal{N}_{j,p}, \sigma_{j,p}^2)$, con $3 \leq i \leq 5$ siendo i el número de gaussianas utilizadas en la mezcla.

$$\bar{\eta}_p = \sum_{j=1}^i w_j (\mathcal{N}_{j,p}, \sigma_{j,p}^2) \quad (1)$$

donde W_j representa el peso de la j -ésima componente Gaussiana y $(\mathcal{N}_{j,p}, \sigma_{j,p}^2)$ es la distribución normal de la j -ésima componente.

Cada cierto intervalo de tiempo T , este fondo es actualizado. Esta actualización se realiza de manera recursiva, adaptando el modelo según un parámetro $\alpha \ll 1$:

$$\mathcal{N}_{j,p} = \alpha I_{p,t} + (1 - \alpha) \mathcal{N}_{p,t-T} \quad (2)$$

$$\sigma_{j,p}^2 = \alpha |I_{p,t} - \mathcal{N}_{p,t}|^2 + (1 - \alpha) \sigma_{p,t-T}^2 \quad (3)$$

donde $\alpha = 1/T$ es una constante de tiempo que determina el factor de cambio entre frames (KaewTraKulPong & Bowden, 2002).

Posteriormente, sobre la imagen de fondo se realizan operaciones de suavizado, umbralizado y morfología matemática para lograr un resultado más estable.

Extracción de objetos en movimiento

Esta etapa se divide en los siguientes 3 pasos.

1. Detección de objetos en movimiento

En este paso se utiliza la imagen de fondo estimada y para cada cuadro I_t (para un tiempo t cualquiera), un pixel es considerado "objeto" si

$$|I_{p,t} - \bar{\mathcal{N}}_p| \geq \epsilon \sigma_p^2 \quad (4)$$

para un umbral ϵ dado. El valor de ϵ otorga al sistema tolerancia al ruido generado en la captura, y es determinado de manera empírica. Finalmente, se obtiene la envolvente convexa de cada objeto, asegurando una completa inclusión del mismo aún si se perdieron píxeles en el procesado morfológico (Gonzalez & Woods, 2002).

2. Etiquetado de objetos

El objetivo de este paso es que cada objeto identificado en la escena obtenga una única etiqueta, que se mantendrá desde el primero hasta el último cuadro donde aparece el objeto en cuestión. Para implementar esto, se computan los centros de masa de los objetos en cada cuadro y se comparan con los calculados en el cuadro previo, mediante la distancia euclídea

$$d_E(C_a, C_b) = \sqrt{(x_a - x_b)^2 + (y_a - y_b)^2} \quad (5)$$

donde $C_a = (x_a, y_a)$ y $C_b = (x_b, y_b)$ son los centros de masa de cuadros sucesivos.

Esta distancia es utilizada para chequear la correspondencia de objetos entre frames sucesivos. Si $d_E \leq \delta$, se considera el mismo objeto y la etiqueta asignada es la misma para ambos. Experimentalmente se fija $\delta = 5$. De esta manera, es posible mantener una misma etiqueta para un objeto ya existente y asignar una diferente a los objetos novedosos en la escena (Pritch et al., 2009).

3. Individualización de objetos

Se procede a la individualización de cada objeto agrupando aquellos cuadros que los incluyen. De esta forma, para cada objeto se obtiene información acerca de: el instante en que aparece en escena, el tiempo total de permanencia, el momento en que sale de escena, la información espacio-temporal de su trayectoria, su tamaño, su textura, sus colores, etc. Finalmente, para cada objeto se genera un nuevo video donde puede verse el comportamiento de éste (de forma aislada) en la escena.

Generación de video sinopsis

La sinopsis puede realizarse de dos formas y para ambos casos se consideran: un fondo fijo y la lista de los objetos que se desea que estén presentes en la escena. La forma más simple de sinopsis genera un video que muestra la actividad en la escena, es decir, que elimina los momentos en que no se detectan movimientos. Una ventaja de este método es que el usuario no debe establecer ningún parámetro adicional.

La segunda opción de sinopsis permite que los objetos seleccionados estén presentes en escena de manera concurrente. Éstos se irán introduciendo en el mismo orden en que fueron apareciendo en el video original. Además, es posible adicionar un pequeño retraso (k cuadros) en cada nueva inserción para evitar que se superpongan los objetos en el cuadro inicial. La cantidad de objetos simultáneos puede ser controlada mediante un parámetro específico.

Adicionalmente, el sistema dispone de parámetros que controlan el tamaño y color de los objetos que aparecen en escena.

EXPERIMENTOS Y RESULTADOS

Para el desarrollo del trabajo se realizaron pruebas sobre distintos videos, de los cuales se tomó uno como ejemplo para este artículo. Los videos considerados para probar el algoritmo poseen una tasa de 30 cuadros por segundo, con resoluciones de 480p (4:3) y 720p (16:9). Estos fueron tomados tanto de la plataforma para compartir videos YouTube, como también de una base de datos privada de videos de vigilancia. La duración de los mismos fluctuaba entre 10 minutos y 1 hora. En la Figura 1 se puede observar una escena del video original, donde aparecen entre 3 a 5 objetos (número promedio de objetos concurrentes).

En la video sinopsis generada se puede sobreimprimir una etiqueta numérica que identifica a cada objeto aparece en escena, indicando el tiempo original de aparición, como puede observarse en la Figura 2.

Finalmente, si se desea obtener algún objeto de tamaño o de un color particular, el método brinda la posibilidad de realizar una segmentación selectiva, obteniendo así una video sinopsis con los objetos deseados.

Las tasas de compresión logradas por el método varían dependiendo de los parámetros seleccionados para ser mostrados en la video sinopsis resultante. En promedio, la tasa de compresión fue de al menos 4:1. Ahora bien, modificando parámetros de visualización, como al mostrar objetos de un color determinado, esta tasa aumentó aproximadamente hasta 50:1.

Un video de ejemplo obtenido con la biblioteca desarrollada se puede encontrar online en <https://youtu.be/gKh53lXPics>.



Figura 1



Figura 2

CONCLUSIONES

En este trabajo se presenta el desarrollo inicial de una biblioteca software de código abierto para la realización de video synopsis a partir de la detección de objetos. La misma se encuentra disponible en <https://github.com/lerker/OpenSourceVS>. Se proveen implementaciones en C++ para las etapas de modelado del fondo, detección y etiquetado de objetos, y fusión de movimientos individuales. La biblioteca presenta, además, una gran cantidad de parámetros que se deben configurar para obtener una synopsis con la información deseada. Estos parámetros están relacionados con condiciones de iluminación, forma de los objetos, color, tamaño, entre otros.

Los resultados preliminares obtenidos son satisfactorios, generando un video de menor duración que condensa la actividad detectada. Asimismo, esta biblioteca permitiría abordar y dar solución a problemas de video vigilancia, control de tránsito, seguridad, compresión, etc., en un sistema abierto y de libre disponibilidad.

REFERENCIAS

- Thounaojam D.**, 2014. et al. A survey on video segmentation. Intelligent Computing, Networking, and Informatics. Springer India, 903-912.
- Fu W., Wang J., Gui L., Lu H., Ma S.**, 2014. Online video synopsis of structured motion. Neurocomputing, vol. 135, pp. 155-162.
- Nam, J., Tewfik, A.**, 1999. Video abstract of video. In Multimedia Signal Processing, 1999 IEEE 3rd Workshop on, IEEE, pp. 117-122.
- Petrovic, N., Jojic, N., Huang, T.**, 2005. Adaptive video fast forward. Multimedia Tools and Applications, 26(3):327-344, Springer.
- Rav-Acha, A., Pritch, Y., Peleg, S.**, 2006. Making a long video short: Dynamic video synopsis. In Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, vol. 1, pp. 435- 441, IEEE.
- Yao, T., Xiao, M., Ma, C., Shen, C., Li, P.**, 2014. Object based video synopsis". In Advanced Research and Technology in Industry Applications (WARTIA), 2014 IEEE Workshop on, pp. 1138-1141, IEEE.
- KaewTraKulPong, P., Bowden, R.**, 2002. An improved adaptive background mixture model for real-time tracking with shadow detection. In Video-based surveillance systems, Springer US, pp. 135-144.
- Gonzalez, R., Woods, R.**, 2002. Digital Image Processing. New Jersey: Prentice Hall.
- Pritch Y., Ratovitch, S., Hendel A., Peleg, S.**, 2009. "Clustered Synopsis of Surveillance video". 6th IEEE Int. Conf. on Advanced video and Signal Based Surveillance, Genoa, Italy.