

Procesamiento de Video Digital: Reconocimiento de Patrones Aplicado a la Detección de Key-Frames

Javier Iparraguirre

Laboratorio de Ciencias de las Imágenes (www.imaglabs.org)
Universidad Nacional del Sur
Avenida Alem 1253, Bahía Blanca, Argentina
j.iparraguirre@computer.org

Claudio Delrieux

Laboratorio de Ciencias de las Imágenes (www.imaglabs.org)
Universidad Nacional del Sur
Avenida Alem 1253, Bahía Blanca, Argentina
cad@uns.edu.ar

RESUMEN

El video digital está convirtiéndose en el medio de comunicación por excelencia, tanto en redes sociales y sitios de video a demanda, como en la inminente llegada de la Televisión Digital. En este trabajo se presenta una metodología de reconocimiento de patrones aplicada a la detección de *Key-Frames*, cuya utilidad futura apunta tanto al *tagging* no supervisado de secuencias de video, como al reconocimiento de contenidos en el mismo. Presentamos los resultados preliminares de una aplicación que por medio del algoritmo SURF permite evaluar diferencias sucesivas entre cuadros, y de esa manera seleccionar key frames candidatos para identificar tomas sucesivas. Esta idea fue testada con video en tiempo real, produciendo resultados satisfactorios.

Palabras clave: Visión por computadoras. Procesamiento de video. Procesamiento paralelo. Reconocimiento de patrones.

CONTEXTO

Este trabajo de investigación se realiza dentro del contexto del Proyecto de Investigación Acreditado *Procesamiento Inteligente de Imágenes* financiado por la SECyT-UNS, dentro del cual el Mg.

Iparraguirre realiza su trabajo de tesis doctoral bajo la dirección del Dr. Delrieux como parte de las actividades del Laboratorio de las Ciencias de las Imágenes de la UNS (www.imaglabs.org). El grupo pertenece al Departamento de Ingeniería Eléctrica y Computadoras de la Universidad Nacional del Sur. También se espera interacción y aportes desde la Facultad Regional Bahía Blanca de la Universidad Tecnológica Nacional.

1. INTRODUCCIÓN

El video en formato digital se ha popularizado en los últimos años [4]. En la vida diaria de la sociedad en general se interactúa con video. El video es una fuente de información en Internet, televisión digital, teléfonos celulares, sistemas de vigilancia y en los hogares. El potencial de la convergencia entre todas estos medios que utilizan actualmente video digital está lejos de ser reconocido y aprovechado.

Uno de los problemas más importantes dentro del procesamiento de video, especialmente en aplicaciones en tiempo real, es el reconocimiento y el *tagging* o etiquetado de contenidos. Si bien el objetivo ideal de tener algoritmos no supervisados

que interpreten el contenido de secuencias de video está en la actualidad fuera de las posibilidades del estado del arte, nuestra opinión es que se pueden diseñar algoritmos autónomos que avancen en esa dirección y produzcan resultados intermedios de importancia, aun en computadoras PC. Por ejemplo, identificar los *key frames* (los frames que separan una toma de otra), identificar el tipo de toma a partir de la dinámica de puntos significativos (si hay objetos en movimiento, zoom, desplazamiento de cámara, etc.). Eventualmente, a partir de algoritmos de reconocimiento de contenidos en imágenes, y del análisis de la dinámica en tomas, se puede llegar a un reconocimiento parcial del contenido de una toma.

Por otra parte, el procesamiento de video es altamente demandante de recursos computacionales [5]. Las computadoras de uso cotidiano apenas pueden procesar unos pocos cuadros por segundo si se utilizan únicamente por medio de lenguajes de programación convencionales. Por lo tanto, otro aspecto no menor de este proyecto consiste en estudiar la paralelización de los algoritmos de procesamiento, y su implementación en diversas arquitecturas multihilo y multicore.

2. DESARROLLO

El trabajo consiste en investigar algoritmos de análisis de video, y su posible implementación en arquitecturas paralelas. La estrategia de trabajo de la primera parte está dividida en tres etapas:

1. **Detección de key-frames.** Para esta etapa empleamos el algoritmo SURF cuadro a cuadro, encontrando un conjunto de puntos distintivos y la descripción de sus características. Entre cuadros sucesivos se evalúa la distancia entre dichos puntos en el espacio de los característicos, y se genera una medida de “distancia acumulada” entre cuadros que toma en cuenta la sumatoria de las distancias entre dichos puntos. Cuando existe una discontinuidad en esta distancia (la cual se puede determinar a partir del promedio de distancias entre cuadros sucesivos anteriores) el algoritmo determina que un cambio importante ha ocurrido entre el cuadro anterior y el actual, y por lo tanto decide que se trata de un key frame.

2. Extracción de contenido de las tomas.

Un análisis de la dinámica de los puntos representativos dentro de cada toma permite determinar la naturaleza de la misma (si hay movimiento de objetos o de cámara, etc.). El análisis de los key frames que demarcan la toma, utilizando software existente de detección de contenidos en imágenes, junto con la naturaleza de la misma, puede a su vez sugerir un contenido posible. En este punto, dependiendo de la naturaleza de la aplicación, es posible hacer un análisis de contenidos inapropiados para facilitar la moderación no supervisada de contenidos. En las redes sociales al estilo de YouTube o Facebook, la moderación de contenidos es semiasistida, muchas veces dependiente de los usuarios, pero en televisión digital va a ser necesario contar con algoritmos no supervisados en tiempo real que realicen este tipo de tareas.

3. **Tagging.** A partir del análisis del contenido de las tomas, se puede etiquetar el video y sugerir posibles nombres para las tomas y key-frames. Por otro lado, se puede hacer un razonamiento basado en casos que contemple la dinámica de videos con contenidos similares y sugiera tags en función de ejemplos etiquetados en forma supervisada (utilizando, por ejemplo, videos de YouTube como feed).

El segundo aspecto de este proyecto, como se mencionara, consiste en investigar el desarrollo de estos algoritmos en arquitecturas de computación paralela. En particular se cuenta con un equipo SUN Ultraspark (detallar características), una supercomputadora basada en GPU que cuenta con 960 cores, y en un futuro se ensamblará un cluster de PCs con GPUs.

3. RESULTADOS OBTENIDOS

En este momento está finalizado el desarrollo de la primera etapa, focalizada en la detección de cuadros de video relevantes. Actualmente disponemos de una aplicación que procesa una secuencia de video utilizando el algoritmo SURF (Speeded Up Robust Features) [1, 2].

En la Figura 1 se muestra una captura de pantalla de la aplicación procesando una secuencia de una cámara en vivo. Alrededor del cuadro 70 se produce un cambio en el video (el usuario mueve la taza), lo cual se refleja en un cambio abrupto en la distancia acumulada entre puntos característicos de los cuadros.

Como muestran estos resultados, la detección de cuadros clave está produciendo los resultados esperados, por lo que el trabajo siguiente consiste en pasar a la segunda etapa (análisis de la dinámica en las tomas).

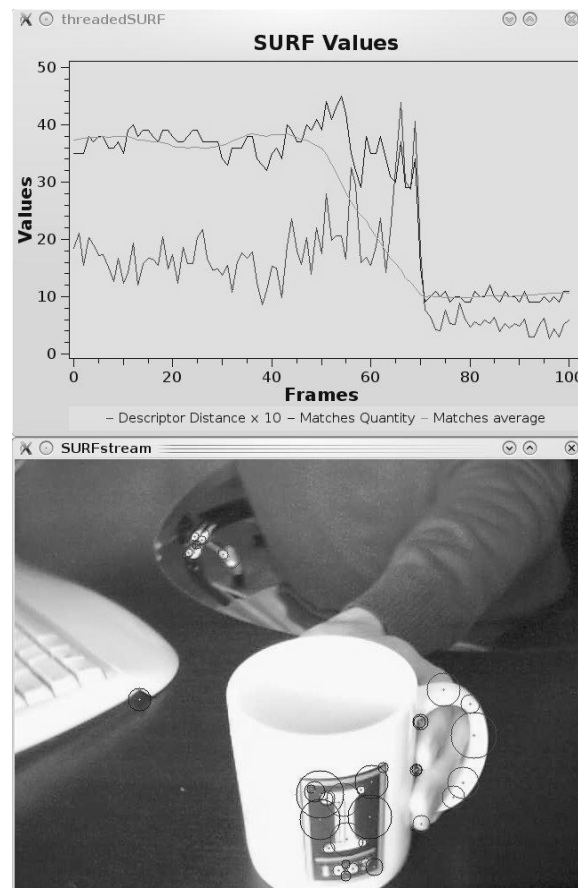


Figura 1: Procesamiento de video mostrado en la parte inferior usando SURF, y la distancia acumulada en los característicos entre cuadros sucesivos (parte superior).

También estamos comenzando a investigar la implementación de estos algoritmos en arquitecturas masivamente paralelas. El espectro de aplicación del desarrollo es muy amplio. Se pueden encontrar aplicaciones tan diversas como realidad aumentada (AR)

[3, 6, 7], interfaz hombre-máquina (HCI) o reconocimientos de patrones [4]. El interés social por el tema es creciente a medida que transcurre el tiempo y esperamos un gran número de aplicaciones prácticas al desarrollo.

4. FORMACIÓN DE RECURSOS HUMANOS

Como parte de las actividades asociadas al proyecto, el Mg. Iparraguirre está realizando su tesis doctoral. Además se están dirigiendo otras 2 tesis doctorales, se desarrollan 2 trabajos finales de carrera, y se realizan cursos de grado y postgrado en Procesamiento de Imágenes. Se está además trabajando en formar un grupo especializado en el procesamiento de imágenes en las dos entidades mencionadas anteriormente.

5. BIBLIOGRAFÍA

[1] H. Bay, T. Tuytelaars, and L. Van Gool, Surf: Speeded up robust features, Lecture notes in computer science, 3951:404, 2006.

[2] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (surf). *Computer Vision and Image Understanding*, 110(3):346 - 359, 2008, Similarity Matching in Computer Vision and Multimedia.

[3] Klein, G. and Murray, D. 2007. Parallel Tracking and Mapping for Small AR Workspaces. In *Proceedings of the 2007 6th IEEE and ACM international Symposium on Mixed and Augmented Reality* (November 13 - 16, 2007). Symposium on Mixed and

Augmented Reality. IEEE Computer Society, Washington, DC, 1-10.

[4] Lew, M. S., Sebe, N., Djeraba, C., and Jain, R. 2006. Content-based multimedia information retrieval: State of the art and challenges. *ACM Trans. Multimedia Comput. Commun. Appl.* 2, 1 (Feb. 2006), 1-19.

[5] Datta, R., Joshi, D., Li, J., and Wang, J. Z. 2008. Image retrieval: Ideas, influences, and trends of the new age. *ACM Comput. Surv.* 40, 2 (Apr. 2008), 1-60.

[6] Takacs, G., Chandrasekhar, V., Gelfand, N., Xiong, Y., Chen, W., Bismpiagiannis, T., Grzeszczuk, R., Pulli, K., and Girod, B. 2008. Outdoors augmented reality on mobile phone using loxel-based visual feature organization. In *Proceeding of the 1st ACM international Conference on Multimedia information Retrieval* (Vancouver, British Columbia, Canada, October 30 - 31, 2008).

[7] Ozuysal, M. Calonder, M. Lepetit, V. Fua, P. Fast Keypoint Recognition Using Random Ferns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32 (3), 448 - 461, March 2010.