

Propuesta de Tesina para la obtención del grado Licenciado en Ciencias de la Computación

9 de diciembre de 2015

Postulante: Martín Reixach, Lucía Norma

Director: Granitto, Pablo

Codirector: Larese, Mónica

1. Situación del postulante

Actualmente me resta aprobar las materias Sistemas Operativos, Inteligencia Artificial y Compiladores, las que planeo completar durante el primer cuatrimestre del próximo año.

En relación a otras actividades académicas, me estoy desempeñando como docente en las materias Programación I y Programación II.

Proyecto dedicar al menos 25 hs. semanales al desarrollo de esta Tesina.

2. Título

Detección de eventos en videos: análisis de partidos de Rugby.

3. Motivación y Objetivo General

El desarrollo de nuevas tecnologías, acompañado de la reducción de costo de las mismas, ha hecho que el acceso a medios para la producción de material audio-visual se haya masificado, cambiando rotundamente la forma en que se trabajaba con videos y generando nuevas herramientas, así como nuevos desafíos y necesidades.

Todo esto ha permitido extender la utilización de cámaras de video a otros ámbitos más allá de la producción de contenidos de entretenimiento, alcanzando distintos objetivos tales como el control de seguridad y el almacenamiento de recuerdos personales, entre otros. Esto genera grandes volúmenes de datos conservados en formato de video. Si bien los mismos pueden contener muchísima información útil, su análisis y clasificación no es sencilla. En ámbitos profesionales, ya sea para análisis de seguridad o contenidos de entretenimiento, la cantidad de material que se produce ha hecho que sea muy costoso y prácticamente inviable analizarlos manualmente. Por esta razón, se busca automatizar el análisis de los mismos.

A lo largo de los últimos años, se han desarrollado distintas técnicas y enfoques para el análisis de videos. Puesto que éstos no son otra cosa que secuencias de imágenes con sonido

incorporado, uno de los enfoques para su análisis se basa en la utilización de los métodos ya desarrollados para imágenes con la incorporación del factor temporal, el cual puede aportar nueva información. Si bien el sonido presente también puede ayudar en el estudio, muchas veces es descartado durante el análisis, basándose éste solamente en los aspectos visuales.

Un método común y que ha resultado ser eficiente para obtener información a partir de imágenes es la utilización de descriptores. Estos descriptores informan acerca de las características visuales de la imagen, describiendo características elementales como la forma, el color, la textura y la ubicación de elementos dentro de la misma. Para extender esta idea al plano de los videos, lo que se hace es desglosar el video en las imágenes que lo componen (frames), para luego analizar por separado cada una de ellas. Con esta información ya disponible se incorpora el factor temporal, siendo éste el eje en la relación entre los distintos frames, permitiendo, por ejemplo, analizar variaciones de un frame a otro.

4. Fundamentos y estado del conocimiento sobre el tema

5. Objetivos específicos

Proponemos desarrollar una herramienta que nos permita identificar *lines*, *scrums* y *momentos de juego* en videos correspondientes a partidos de Rugby.

Para lograrlo nos basaremos en el método de Bag-of-Words junto a descriptores de imágenes y videos (en este trabajo el audio del video no será tenido en cuenta).

También se analizará la *performance* de los métodos de detección usados y se compararán los resultados obtenidos con los alcanzados utilizando otros métodos.

6. Metodología y Plan de Trabajo

Para alcanzar los objetivos propuestos proponemos las siguientes tareas (programa tentativo de trabajo):

- Estudio de los métodos más usados para análisis de videos e imágenes. *6 semanas*
- Implementación de los algoritmos necesarios. *4 semanas*
La implementación se hará de manera progresiva, incorporando las herramientas que sean necesarias para obtener información más clara.
- Evaluación del funcionamiento de los métodos sobre los videos de nuestra base de datos y ajuste de los parámetros según sea necesario. *4 semanas*
- Síntesis de los resultados obtenidos y escritura de la tesina. *6 semanas*

El trabajo se realizará durante aproximadamente 5 meses con una dedicación de más de 25 hs. semanales.

Referencias

- [1] A. Alahi, R. Ortiz, and P. Vandergheynst. Freak: Fast retina keypoint. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 510–517. IEEE, 2012.

- [2] M. Chen and A. Hauptmann. Mosift: Recognizing human actions in surveillance videos. research showcase. *Computer Science Department, School of Computer Science, Carnegie Mellon University*, 2009.
- [3] O. Deniz, I. Serrano, G. Bueno, and T. Kim. Fast violence detection in video. In *The 9th International Conference on Computer Vision Theory and Applications (VISAPP), Lisbon, Portugal*, 2014.
- [4] O. Kliper-Gross, Y. Gurovich, T. Hassner, and L. Wolf. Motion interchange patterns for action recognition in unconstrained videos. In *Computer Vision, ECCV 2012*, pages 256–269. Springer, 2012.
- [5] I. Laptev. On space-time interest points. *International Journal of Computer Vision*, 64(2-3):107–123, 2005.
- [6] S. Leutenegger, M. Chli, and R. Y. Siegwart. Brisk: Binary robust invariant scalable keypoints. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2548–2555. IEEE, 2011.
- [7] E. B. Nieves, O. D. Suarez, G. B. García, and R. Sukthankar. Violence detection in video using computer vision techniques. In *Computer Analysis of Images and Patterns*, pages 332–339. Springer, 2011.
- [8] C. Whiten, R. Laganieri, and G.-A. Bilodeau. Efficient action recognition with mofreak. In *Computer and Robot Vision (CRV), 2013 IEEE International Conference on*, pages 319–325. IEEE, 2013.