An Architecture for Social Media Summarisation

Zbigniew Zdziarski, Joe Mitchell, Pierre Houdyer, Dave Johnson
Tapastreet Ltd, Ireland
{zbigniew, joe, pierre, dave}@tapastreet.com

Cyril Bourgès & Rozenn Dahyot
Trinity College Dublin, Ireland
{bourgesc, Rozenn.Dahyot}@tcd.ie

Abstract

Social media traffic and mobile usage is growing at an accelerating rate, and the amount of media that is being uploaded on social media sites (such as Twitter, Facebook and Instagram) is also increasing. The consortium GRAISearch aims at developing tools to merge, to visualise and to present this wealth of data in a comprehensive, compact and user-friendly way. This poster will present a work-in-progress architecture for such a purpose - to provide users with a central point of access to media from the largest social media sites.

Keywords: Social Media, Web Harvesting, Video Summarisation, Visual Saliency

The European Project GRAISearch (Use of Graphics Rendering and Artificial Intelligence for Improved Mobile Search Capabilities, FP7-PEOPLE-2013-IAPP (612334), 2014-18) is a research collaboration between two universities (Trinity College Dublin, Ireland, and INSA Lyon, France) and the company Tapastreet Ltd (see mobile app at `http://tapastreet.com/`). It aims at providing enhanced visualisation tools for visual content available on social media and an architecture for social media summarisation. Tapastreet has a location based social media search engine platform that, in its current form, returns geo-located video and image media from major social networks for any location and any topic (#hashtags) anywhere in the world. The current platform deals well with images on social media but videos are yet not well tackled. Several challenges exist for videos on social platforms. First, they are too large to all be downloaded when browsing on mobile devices and therefore need to be summarised very efficiently. Second, media on social platforms consist mainly of very diverse amateur recordings with little or no editing rules that also contain many artefacts that alter their quality, such as low lights and motion shakiness when the recording device is hand-held.

Harvesting Social media. Using social network APIs, a Ruby script is used to download all media using a user-defined query (hashtags, GPS location), and links to these images and videos are stored along with their description in JSON format (keywords, GPS location, creation date, etc.) on our server. An image-processing pipeline is currently under development for automatically creating video summaries.

Video Summarisation. A lot of research has been devoted to creating video summarisations (a.k.a. video abstractions) [Truong and Venkatesh, 2007] and many algorithms have been proposed. For example, Zhang et al. [Zhang et al., 2003] & Kim and Hwang [Kim and Hwang, 2002] suggest processing a video sequentially and marking keyframes as those that are significantly different from previously extracted keyframes. A more computationally demanding method has been proposed by Gibson et al. [Gibson et al., 2002] and Yu et al. [Yu et al., 2004]. They employ a clustering technique where video frames are treated as points in a feature space (e.g. colour histogram) and representative points from each cluster are selected as keyframes of

the video. The keyframes extracted by these methods can then be presented to the user as a summary of the video.
For the GRAISearch project, a real-time system is required, hence video summarisation needs to be very fast as well as informative (i.e. representative of the video) and very small in storage size (due to the limitations of mobile devices and wireless networks [Liu et al., 2014]). Several techniques are currently being tested using information theory (e.g. measure of entropy) to select the most diverse frames in a stream, and visual saliency algorithms [Hou and Zhang, 2007] to assist in detecting salient regions in frames and hence improve the keyframe extraction process. A further extension to this project is to develop video summaries suitable for 3D screens and for this 3D visual saliency algorithms [Zdziarski and Dahyot, 2014] will also be investigated.

Examples of scenarios. Tapastreet Ltd currently have a number of outside bodies actively using their app. The Danish Football Association uses the app as a fan engagement tool. Whenever the Danish football team plays, the association channels all images onto their website that were taken at a particular stadium (specified by GPS location) and/or tagged with appropriate hashtags for users to view in real-time. The MET office in the UK uses the app to measure the impact of weather events on human activity. When it knows that a significant weather event is imminent for a particular area, it will mine all media from that area before, during and after the event, allowing the study of the impact of climate change on human behaviour.

References

[Gibson et al., 2002] Gibson, D., Campbell, N., and Thomas, B. (2002). Visual abstraction of wildlife footage using gaussian mixture models and the minimum description length criterion. In *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, volume 2, pages 814–817 vol.2.

[Hou and Zhang, 2007] Hou, X. and Zhang, L. (2007). Saliency detection: A spectral residual approach. In *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 1–8.

[Kim and Hwang, 2002] Kim, C. and Hwang, J.-N. (2002). Object-based video abstraction for video surveillance systems. *Circuits and Systems for Video Technology, IEEE Transactions on*, 12(12):1128–1138.

[Liu et al., 2014] Liu, Y., Wang, S., and Dey, S. (2014). Content-aware modeling and enhancing user experience in cloud mobile rendering and streaming. *Emerging and Selected Topics in Circuits and Systems, IEEE Journal on*, 4(1):43–56.

[Truong and Venkatesh, 2007] Truong, B. T. and Venkatesh, S. (2007). Video abstraction: A systematic review and classification. *ACM Trans. Multimedia Comput. Commun. Appl.*, 3(1).

[Yu et al., 2004] Yu, X.-D., Wang, L., Tian, Q., and Xue, P. (2004). Multilevel video representation with application to keyframe extraction. In *Multimedia Modelling Conference, 2004. Proceedings. 10th International*, pages 117–123.

[Zdziarski and Dahyot, 2014] Zdziarski, Z. and Dahyot, R. (2014). Extension of gbvs to 3d media. In *Signal Processing and Communications Applications Conference (SIU), 2014 22nd*, pages 2296–2300.

[Zhang et al., 2003] Zhang, X.-D., Liu, T.-Y., Lo, K.-T., and Feng, J. (2003). Dynamic selection and effective compression of key frames for video abstraction. *Pattern Recognition Letters*, 24(9-10):1523 – 1532.