# The role of multimedia in archiving community memories

Jonathon S. Hare, David P. Dupplaw, Wendy Hall, Paul H. Lewis, and Kirk Martinez

Electronics and Computer Science,
University of Southampton, Southampton, UK
{jsh2,dpd,wh,phl,km}@ecs.soton.ac.uk

**Abstract.** The data contained on the web and social web is inherently multimedia; consisting of a mix of textual, visual and audio modalities. Community memories embodied on the web and social web contain a rich mixture of data from these modalities. This paper explores some uses for the automatic analysis of multimedia data within the context of the archival and post-archival analysis of community memories on the web and social web.

## 1 Introduction

Community memories embodied on the web and in social media are inherently multimedia in nature, and contain data and information in many different audio, visual and textual forms. From the perspective of archiving and preservation, taking these modalities into account is very important. As we move towards more intelligent archiving, and more intelligent post-crawl analysis of the data hidden within the archive, the ability to leverage the vast array of different data modalities becomes even more crucial.

This paper discusses some of the needs and opportunities for both analysing the multimedia data within an archive of community memories, and exploiting the data during the creation of the archive. Practical embodiments of the discussed techniques within the context of the EU ARCOMEM project[1] are also described.

## 2 Use cases for multimedia analysis in archiving community memories

There are numerous applications for multimedia analysis within a community memory archiving scenario. As described in the following use cases, multimedia analytics has potential applications in both post-archival analysis as well as a role in helping to identify what should be crawled.

[1] http://www.arcomem.eu

### 2.1 Use case 1: Aggregating social commentary

Media is often duplicated on the web. For example, on YouTube it is common to find multiple versions of the same video, often with minor changes (i.e. different sound tracks or modified visual appearance — it's common for the original broadcaster's logo to be changed in a bootleg copy). On Twitter, the same effect is observed, and it's common for the same/similar media object to be shared and re-posted (at a different URL) many times [3] (see the second use-case below). In both these examples, the communities sharing and providing commentary on individual instances of a particular media item tend not to overlap.

From the perspective of archiving community memories, it is important to find and capture the relationships between near-duplicate media objects as this allows the social context and commentary to be aggregated, building a much fuller and potentially more diverse picture of the community associated with the media object. An example of this is shown in Figure 1.

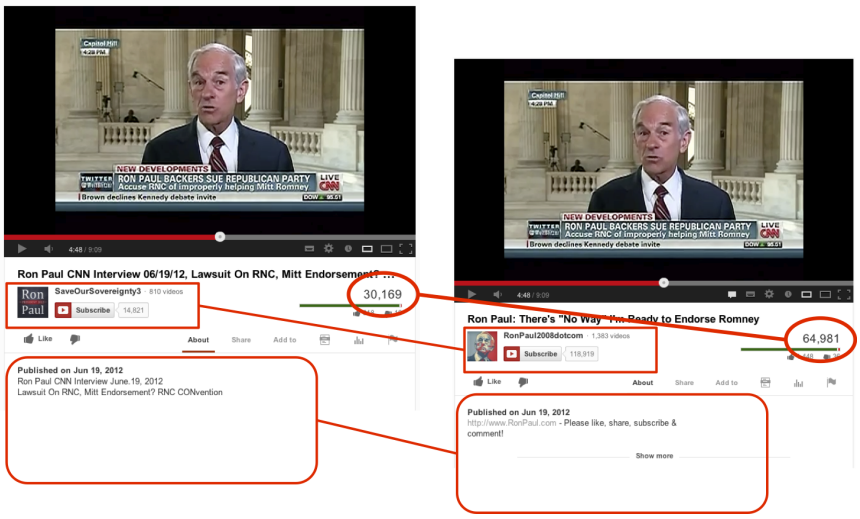


**Fig. 1.** Aggregating statistics and commentary across duplicate videos.

From a practical perspective, it can be very difficult or even impossible to determine whether two media items are duplicated by looking at the metadata alone. However, modern image analysis and indexing techniques allow duplicates to be detected efficiently [3, 2].

Beyond the aggregation of social context, the detection of duplicates also has another potential use in archiving as it can allow the overall size of the archive to be reduced without losing information as it is often unnecessary to store multiple copies of content that is identical. It may also be preferable to find and store only the highest quality copy or perhaps the smallest in size for certain archives.

### 2.2 Use case 2: The temporal pulse of social multimedia

The detection of interesting trends and topics within streams of communication within a social network is clearly of importance to the problem of deciding what needs to be included within a community memory archive. Traditional approaches to trend and topic detection are based on the analysis of textual content within the communication stream. However, it is possible to analyse the other data modalities within the stream to the same ends.

An example of this is our recent work on the detection of near-duplicate images posted on Twitter [3]. Within this work we developed the tools required to analyse images posted on Twitter in near-realtime in order to detect clusters of very similar images that are being tweeted about within a window of time. These clusters represent the key influential images that are currently trending. An example, is illustrated in Figure 2 which shows a Korean pop-group that was trending in January 2013.


**Fig. 2.** Visualisation of trends in Twitter detected by looking for near-duplicate images within a time window.

### 2.3 Use case 3: Detecting media about individuals, organisations and places

Modern web archiving techniques, such as those being created by the ARCOMEM project, are often centred around the main entities (people, places and organisations) that described the crawl domain. Image analysis can be used to help select relevant multimedia documents by looking for the presence of specific entities in a visual sense.

For people entities, modern advancements in dynamic face recognition can be applied to learn what a person looks like from a small set of images (typically collected through web-search if the person is well known) (e.g. [5]). Organisations

can be detected by looking for the presence of their corporate logo in images; there are many different ways in which logo recognition can be achieved, but slight variants of the techniques used in the first two use-cases are popular. Finally, place detection is an emerging research area; if the depicted location is visually unique then it may be solved by matching against a large corpus of geo-located images [2]. For other images, intelligent guesses can be made using a combination of visual features and any available metadata (such as the caption text or any tags) (e.g. [4]).

### 2.4 Use case 4: Measuring opinion and sentiment

The analysis of opinions and sentiment within a collection of community memories is often an important requirement of the end users of the archive. Specific examples include sentiment-based search (e.g. 'find me positive media about $X$'), and the determination of the key influential media objects (c.f. use case 2).

Images are often used to illustrate the opinions expressed by the text of a particular article. By themselves, images also have the ability to convey and elicit opinions, emotions and sentiments. In order to investigate how images are used in the opinion formation process, we have been developing tools that (a) allow the reuse of images within an archive to be explored with respect to diverse time and opinion axes; and (b), allow in-depth analysis of specific elements (in particular, the presence and expression of human faces) within an image to be used to quantify opinion and sentiment.

Whilst predictions of the opinions and sentiment of visual content can be made by considering the visual content alone, a richer approach is to consider both the image and the context in which it appears. State-of-the-art research on the sentiment analysis of images [9, 8, 10] has already begun to explore how the analysis of textual content and the analysis of visual content can complement each other. In particular, in one experiment, the results of sentiment classification of images based on the sentiment scores of Flickr tags was combined with a sentiment classifier that was based purely on image features. The performance of the multimodal fusing of the classifiers outperformed either classifier alone [9]. Aspects of attractiveness [6] and privacy [11] (e.g. is this a very private photo being shown in a public place), which aid sentiment classification, are also being explored.

## 3 Integrated multimedia analysis in the ARCOMEM system

Within the ARCOMEM project, together with our project partners, we are building software tools for crawling and analysing samples of web and social-web data. The ARCOMEM software is designed to work with small (tens/hundreds of gigabytes) to medium (multi-terabyte) web datasets.

Fundamentally, the ARCOMEM software has four goals; firstly it provides a way to intelligently harvest data from the web and social web around specific entities, topics, and events. Secondly, it provides a scalable and extensible

platform for analysing harvested datasets, and includes modules for state-of-the-art multimodal (textual, visual and audio) content analysis. Thirdly, it exposes the results of the analysis in the form of a knowledge base that is interlinked with standard linked-data resources and accessible using standard semantic technologies. Finally the software provides the ability to export the harvested and analysed dataset in standardised formats for preservation and exchange.

In terms of multimedia analysis and the use cases given in Section 2, the ARCOMEM software contains a module for efficiently finding near-duplicate image content in images and videos which address the first and second use cases. Experimental support for entity detection is also included as a module in the system (use case 3). The tools for supporting sentiment, privacy and attractiveness classification of images will be included in the software at a later date.

At present, the multimedia analysis modules are implemented as "offline" processes within the ARCOMEM system [7, 1]; that is, they can be run on an archive after it has been crawled. They do not at present direct the crawl in any way. That functionality is provided by fast text analysis modules.

## 4 Conclusions and Outlook

This paper has discussed some of the opportunities created by combining modern multimedia analysis techniques with tools for crawling and performing analysis on archives of community memories. The paper has also described how some of these techniques are being made available within the tools developed by the ARCOMEM project.

Looking ahead to the future, there are a number of research areas in which the use of multimedia analysis could be expanded with respect to the archiving of community memories. Two specific areas for further exploration are outlined below:

1. **Image-entity guided crawling.** At the moment within the ARCOMEM tools, the visual entity tools are applied to the archive after it has been created. However, visual entities could potentially be used to directly influence the crawl process; for example, if a crawl was specified to look at topics surrounding the Olympic games, then the crawl specification could contain images of the Olympic rings logo, and the visual analytics tools could be used to detect the presence of this logo in images. Pages with embedded images with the logo and the outlinks of these pages could then be given higher priority by the crawler.
2. **Image-entity co-reference resolution in multilingual corpora.** Image content is inevitably reused across different documents; often the image will have been scaled or cropped as it is used in different documents. Our tools for detecting this kind of reuse are now quite robust and are capable of providing coreference resolution of the images and the entities they depict. This has many practical uses; for example, it could be used to link documents in different languages as being related, even though we may not have NLP

tools for the languages in question. In turn this coreference information could be used to help guide the crawler to new relevant content.

## Acknowledgements

This work was funded by the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement 270239 (ARCOMEM).

## References

1. Hare, J., Dupplaw, D., Hall, W., Lewis, P.H., Martinez, K.: Building a multimedia web observatory platform. In: Building Web Observatories (May 2013), http://eprints.soton.ac.uk/352465/
2. Hare, J., Samangooei, S., Dupplaw, D., Lewis, P.: Imageterrier: An extensible platform for scalable high-performance image retrieval. In: The ACM International Conference on Multimedia Retrieval (ICMR 2012) (2012)
3. Hare, J.S., Samangooei, S., Dupplaw, D.P., Lewis, P.H.: Twitter's visual pulse. In: Proceedings of the 3rd ACM conference on International conference on multimedia retrieval. pp. 297–298. ICMR '13, ACM, New York, NY, USA (2013), http://doi.acm.org/10.1145/2461466.2461514
4. Hays, J., Efros, A.A.: im2gps: estimating geographic information from a single image. In: Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (2008)
5. Parkhi, O.M., Vedaldi, A., Zisserman, A.: On-the-fly specific person retrieval. In: International Workshop on Image Analysis for Multimedia Interactive Services. IEEE (2012)
6. Pedro, J.S., Siersdorfer, S.: Ranking and classifying attractiveness of photos in folksonomies. In: 18th International World Wide Web Conference. pp. 771–771 (April 2009), http://www2009.eprints.org/78/
7. Risse, T., Peters, W., Senellart, P.: Social and semantic driven web harvesting. In: Building Web Observatories. Paris, France (May 2013)
8. Schmidt, S., Stock, W.G.: Collective indexing of emotions in images. a study in emotional information retrieval. Journal of the American Society for Information Science and Technology 60(5), 863–876 (2009), http://dx.doi.org/10.1002/asi.21043
9. Siersdorfer, S., Hare, J., Minack, E., Deng, F.: Analyzing and predicting sentiment of images on the social web. In: ACM Multimedia 2010. pp. 715–718. ACM (October 2010), http://eprints.ecs.soton.ac.uk/21670/
10. Yanulevskaya, V., van Gemert, J.C., Roth, K., Herbold, A.K., Sebe, N., Geusebroek, J.M.: Emotional valence categorization using holistic image features. In: IEEE International Conference on Image Processing (2008), http://www.science.uva.nl/research/publications/2008/YanulevskayaICIP2008
11. Zerr, S., Siersdorfer, S., Hare, J., Demidova, E.: Privacy-aware image classification and search. In: Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval. pp. 35–44. SIGIR '12, ACM, New York, NY, USA (2012), http://doi.acm.org/10.1145/2348283.2348292