Visual-Based Transmedia Events Detection

Alexis Joly, Julien Champ
INRIA ZENITH
Montpellier, France

Pierre Letessier
INRIA / INA

Nicolas Hervé, Olivier Buisson, Marie-Luce Viaud
INA
Bry-sur-Marne, France

ABSTRACT

This paper presents a visual-based media event detection system based on the automatic discovery of the most circulated images across the main news media (news websites, press agencies, TV news and newspapers). Its main originality is to rely on the transmedia contextual information to denoise the raw visual detections and consequently focus on the most salient transmedia events.

Categories and Subject Descriptors: H.3 [Information Storage And Retrieval]: Miscellaneous

Keywords: Event Detection, Transmedia, mining

1. INTRODUCTION

Automatically detecting, tracking and indexing media events is crucial for structuring and searching news media as well as huge audio-visual archives. Text-based search and mining methods are definitely a powerful tool in this regard but in this work we are rather interested in **visual** news contents, i.e. images illustrating the news delivered to the general public. The visual content of a media event is actually an integral part of its semantic and even sometimes the main reason of its existence or media impact (the Jyllands-Posten Muhammad cartoons are for instance the main reason of the worldwide media controversy itself).

However, one of the limitation of current visual search and mining methods, is that they suffer from a high level of noise, even when they attempt to stick on accurate visual matches such as copies, near-duplicates or rigid objects [4, 5]. Frequent visual items are actually likely to depict undesired contents from the media event mining perspective (commercials, inserted texts, banners, etc.). The main claims of our approach is that (i) many of these false positives can be filtered by considering only **transmedia visual matches** and (ii) most salient media events can be easily detected by looking at the **transmedia contextual information**. As an example, a press agency picture that is circulated over several websites, TV channels and newspapers is more likely to depict a salient media event than a highly frequent image on a single news media.

2. NEWS CRAWLING AND INDEXING

Four categories of *news media* are crawled and indexed by our system with more or less streams in each category:

Category 1 - Press agencies: the whole stream of AFP's multimedia content is indexed in our system. AFP is the oldest press agency in the world and one of the three largest. Each *multimedia news report* includes at least one image, a core text and a set of metadata.

Category 2 - News websites: about 1,500 French websites are crawled from RSS feeds and indexed in our system. Targeted websites are *news* pure players (12), radios (14), TVs (17), press (71), blogs (339), political actors such as senators, deputies, presidential candidates (646), trades unions (19), political parties (263) and institutions (62).

Category 3 - Newspapers: 7 national newspapers are collected numerically and indexed in our system (La Croix, Liberation, Le Monde, 20 Minutes, Direct Matin, les Echos, le canard enchaine).

Category 4 - TV news: News & magazines from 10 French TV channels are collected and indexed in our system: 5 public broadcasting, 3 generalist private channels and 2 news channels.

To ensure a homogeneous integration of these heterogeneous contents, each news document follows a specific workflow depending on its media support. The first step collects native streams metadata and imports additional resources (subtitles, manually annotated notice from TV and Radio databases). Then the content is segmented and structured (title, summary, paragraphs, author, date). Hyperlinks, videos, images are extracted and Media Metadata Files gathering media description (caption, size, duration, author...) are generated. TV and Radio streams are processed by a speech-to-text software. Finally, textual documents are processed by a named entity recognition tool (peoples, places and organizations) and a noun phrases extractor. As discussed further, these enriched metadata are used on top of our visual-based media events detector to automatically build the description of the related media events in the GUI.

All pictures encountered in the streams are considered for content-based visual indexing. A weekly visual index is therefore constructed at the end of each week. For the last category (TV streams), one visual frame is extracted every second and indexed. Visual indexes are built according to the following steps: SIFT features are first extracted from all images and then converted to space-efficient and time-efficient binary hash codes using RMMH [3]. 256 hash functions are generated with this family and used to convert SIFT features into 256-bits hash codes. Hash codes are stored in two ways (with full replication): (i) a simple document-oriented storage allowing to get efficiently the hash codes of a given query document (ii) a hash table using the first 24 bits of the hash codes as keys and the full hash code as value. This allows performing efficient approximate KNN-search of any query hash code using the a posteriori multi-probe strategy of [1] and a Hamming distance as similarity metric during the filtering step.

MM'12, October 29–November 2, 2012, Nara, Japan.
ACM 978-1-4503-1089-5/12/10.

3. TRANSMEDIA VISUAL MINING

Precise Visual Search: The search of a query image in the visual index of a week is done using the method first introduced in [2]. It allows the discovery of different kinds of trust-full visual matches such as copies, part-based copies, near-duplicates, and even smaller rigid visual objects such as logos, trademarks, buildings, faces of the same person with the same pose, manufactured objects, etc. The basic principle is the following: approximate KNN's of each local feature of the query image are searched in the index and the resulting candidate visual matches are filtered with a RANSAC-like geometry consistency checking. Retrieved images are finally further filtered by the a contrario thresholding method described in [2] and that allows a good control of the false positives rate.

Transmedia Search strategy: In order to drastically reduce the search load, TV frames are not searched at all (but might be still retrieved in the index by the queries of the other 3 media categories). Since we are mostly interested in transmedia visual matches, this is actually a useful trick that does not prevent to discover pictures or objects instantiated in at least two media categories (including the TV category). All images of the three other media categories are searched in the visual index of the week they belong as well as in the two neighboring weeks.

Clustering & Transmedia filtering: The matching graphs produced by the previous step for a given period of interest is then processed by the MCL algorithm [6] in order to group images in relevant clusters. Finally, we rank the discovered clusters according to a diversity score in the number of media supports (i.e. using the transmedia *context*). More practically, we first assigned off-line a prefixed weight to any of the considered media. This is done by first equally weighting the media categories: a mass of 0.25 is assigned to each of the 4 categories. The mass of a given media category is then distributed uniformly between all the media streams or programs belonging to it. The mass of AFP stream for instance remains equal to 0.25 whereas the mass of each of the TV channels is 10 times lower. Now, the diversity score of a discovered cluster is computed by summing the weights of the media composing it. Only the top-k clusters according to this score are finally kept as interesting visual media events. Note that the key issue of this transmedia filtering is to effectively remove a lot of noisy clusters such as ads, channels settings, newspapers headers, banners, etc. Visual matches arising across all media categories are actually much more unlikely to be false positives than mono-media visual matches. Figure 1 illustrates a noisy cluster that was pruned by the transmedia contextual reranking and filtering and, on the other side, the top-3 clusters selected by our method for a given week. It well shows that the clusters can have more or less visual diversity in the detected patterns (from small objects to duplicated images). But still, they all correspond to very salient media events for that week.

4. NAVIGATION IN NEWS IMAGES

A front-end GUI built on top of the system allows browsing the *image of the day*, the *image of the week* or the *image of the month* thanks to an interactive timeline. Snapshots of the GUI are presented in Figure 4. The image of each period (day, week or month) and the associated description and keywords were computed as follows: (i) our clustering & contextual reranking method is applied to the visual matching graph restricted to the respective period (ii) only the top-1 cluster (according to the diversity score) is kept (iii) the most frequent keywords in the cluster's metadata are computed (iv) we then choose the image of the cluster having the highest number of occurrences of these keywords and use the corresponding description as event description. Figure 4 presents the result obtained for three successive weeks from the 3rd of October 2011 to the 24th of October 2011.

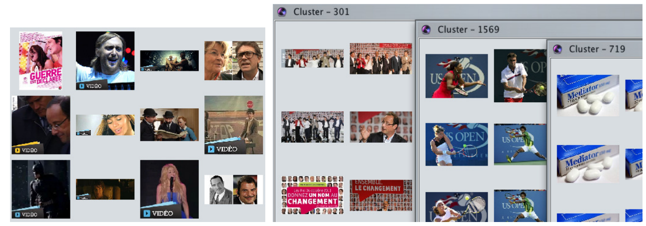

Figure 1: (left) A pruned noisy cluster, (right) top-3 selected clusters for the first week of Sept 2011 (presidential primary, USOpen, mediator scandal)

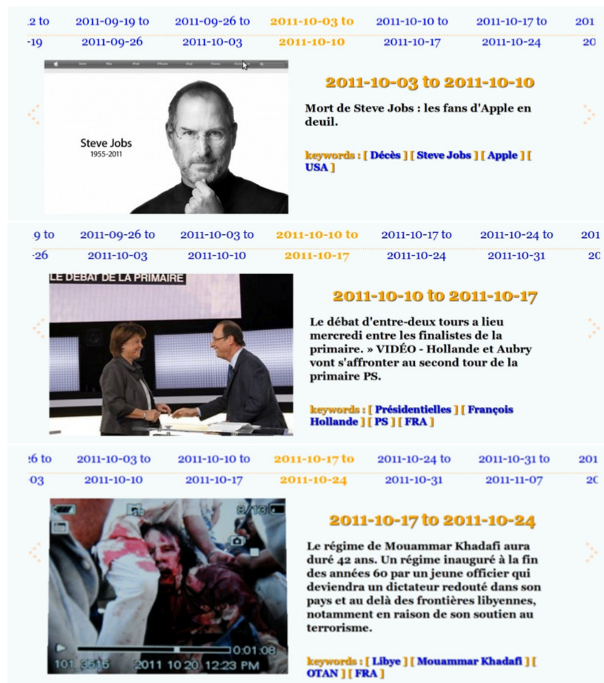

Figure 2: Images of the weeks

5. REFERENCES

[1] A. Joly and O. Buisson. A posteriori multi-probe locality sensitive hashing. In *proc. of ACM MM*, pages 209–218, 2008.

[2] A. Joly and O. Buisson. Logo retrieval with a contrario visual query expansion. In *proc. of ACM MM*, pages 581–584, 2009.

[3] A. Joly and O. Buisson. Random maximum margin hashing. In *CVPR*, Colorado springs, United States, June 2011. IEEE.

[4] A. Joly, O. Buisson, and C. Frélicot. Content-based copy retrieval using distortion-based probabilistic similarity search. *IEEE Transactions on Multimedia*, 9(2):293–306, 2007.

[5] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Lost in quantization: Improving particular object retrieval in large scale image databases. In *proc. of CVPR*, 2008.

[6] S. Van Dongen. A cluster algorithm for graphs. *Report-Information systems*, (10):1–40, 2000.