Social Event Detection at MediaEval 2011: Challenges, Dataset and Evaluation

Symeon Papadopoulos
CERTH / Informatics and Telematics Institute
57001 Thermi, Greece
papadop@iti.gr

Raphael Troncy
EURECOM
Sophia Antipolis, France
raphael.troncy@eurecom.fr

Vasileios Mezaris
CERTH / Informatics and Telematics Institute
57001 Thermi, Greece
bmezaris@iti.gr

Benoit Huet
EURECOM
Sophia Antipolis, France
benoit.huet@eurecom.fr

Ioannis Kompatsiaris
CERTH / Informatics and Telematics Institute
57001 Thermi, Greece
ikom@iti.gr

ABSTRACT

This paper provides an overview of the Social Event Detection (SED) task, which is organized as part of the MediaEval 2011 benchmarking activity. With the convergence between social networking and multimedia creation and distribution being experienced on a regular basis by hundreds of millions of people worldwide, this task examines how new or state of the art techniques can cope with the need for detecting social events by automatically analyzing the social multimedia content. This paper discusses the challenges set as part of the SED Task, the dataset that was provided to the task participants, and the process of evaluating the submissions.

Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]: Information Search and Retrieval

General Terms

Experimentation

1. INTRODUCTION

The modeling, detection, and processing of events is an area that has started to receive considerable attention by the multimedia community [2]. The Social Event Detection (SED) task of MediaEval 2011 requires participants to discover events and detect media items that are related to a specific social event of a given event class. By social events, we mean that the events are planned by people, attended by people and that the social media are captured by people. A lot of multimedia content on the Internet was captured during such an event or is otherwise related to events. However, this content is often scattered, i.e., disassociated from the related events. This, together with the observation that humans often think in terms of events, generate the need for automatically establishing the event-media associations that will allow multimedia browsing and search in a way that is more natural to the users.

MediaEval 2011 Workshop, September 1-2, 2011, Pisa, Italy

2. CHALLENGES

The SED task is composed of two challenges with a common test dataset of images with their metadata (time-stamps, tags, geotags for a small subset of them). Participants were invited to submit results to either one of the challenges, or to both of them. In both cases, the image metadata that can be used by the participants for completing this task are only those provided to them as an XML file. The use of additional information (e.g. geotags) that may be available on the Internet for a given image of the dataset is not permitted. However, external resources (such as Wordnet, Wikipedia, or even visual concept detectors trained on external collections) can be employed, provided that they do not relate to specific images of the test dataset (or any images given for specifying the sought events), and that their development and use did not benefit from any knowledge of the task's dataset and challenge definitions.

2.1 Challenge 1

The first challenge reads: **Find all soccer events taking place in Barcelona (Spain) and Rome (Italy) in the test collection. For each event provide all photos associated with it.**

Soccer events, for the purpose of this task, are soccer games and social events centered around soccer such as the celebration of winning a cup. In contrast, a single person playing with a soccer ball out in the street is not a soccer event under the task's definition.

Finding the events, in this task, does not mean finding some textual descriptions or metadata of the events. What we are looking for is a set of photo clusters, each cluster comprising only photos associated with a single soccer event (thus, each cluster defining a retrieved soccer event). The "photos associated with a soccer event" that we are looking for are all photos of the test collection that directly relate (in content, and also in terms of place/time) with the event of interest. E.g., photos of game x being played, photos of fans inside the stadium during/a bit before/a bit after game x, photos of fans leaving the stadium after the end of game x, are all related to the "game x" soccer event. On the contrary, photos that miss the above relations to an actual soccer event (e.g. a photo showing part of the stadium where

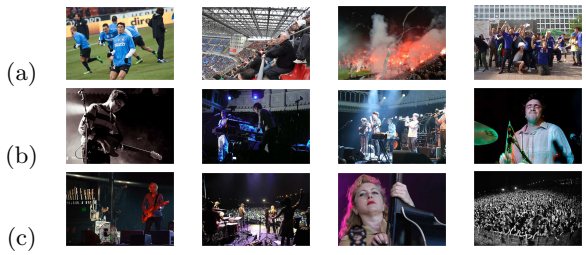

Figure 1: Example images of (a) soccer events, (b) events in Paradiso, (c) events in Parc del Forum.

the fans gather, with no fans visible or otherwise any relation to a specific game), are not considered as relevant.

If all images were properly tagged and correctly geotagged and time-stamped, this would be a trivial task. But, since most images are not geotagged (both in our testset and on the Internet), participants need to also consider tag and/or visual information for finding the most complete set of relevant events and images. As a required (baseline) run, the participants are asked to use any combination of the available image metadata they see fit, but no visual information, for finding the relevant events and images. The use of visual information in addition to the various provided image metadata is encouraged in subsequent runs. Examples of images that are relevant to soccer events are given in Figure 1(a).

2.2 Challenge 2

The second challenge reads: **Find all events that took place in May 2009 in the venue named Paradiso (in Amsterdam, NL) and in the Parc del Forum (in Barcelona, Spain). For each event provide all photos associated with it.**

For both these venues, more than one event took place in May 2009. We consider that multiple bands playing the same evening are not distinct events, but a lineup of multiple artists (i.e. we consider that two different events cannot happen the same day at the same location). Some events (e.g. a festival) can last several days with a lineup of artists and will be considered as a single event.

What we are looking for is again a set of photo clusters, each cluster comprising only photos associated with a single event. For specifying these events, besides the venue names, some exemplary images are provided. These, however, do not have time-stamps. Similarly to the first challenge, participants may need to consider different kinds of information for finding the most complete set of relevant images. A baseline run that uses no visual information is required, and the use of visual information in addition to the various image metadata is encouraged in subsequent runs. Examples of relevant images for the Paradiso and Parc del Forum venues are given in Figure 1(b)-(c).

3. DATASET

A collection of 73.645 photos was created by issuing appropriate queries to the Flickr web service through its web-based API. The collected photos represent the complete set of geotagged photos that were available for five different cities (i.e., Amsterdam, Barcelona, London, Paris and Rome, based on the geotags) and were taken in May 2009, further augmented with a few non-geotagged photos for the same cities and time period [3]. However, before providing the XML photo metadata archive (including any tags, geotags, time-stamps, etc.) to the task participants, the geotags were removed for 80% of the photos in the collection (randomly selected). This was done for simulating the frequent lack of geotags in photo collections on the Internet (including the Flickr collection). The dataset and the ground truth will be made publicly available from the MediaEval website.

4. EVALUATION

The evaluation of the submissions to the SED task is performed with the use of the ground truth EventMedia associations [3]. As an aid, the cluster-based event detection framework of [1] was employed in generating this ground truth. Two evaluation measures are used:

- Harmonic mean (F-score) of Precision and Recall for the retrieved images. This measures only the goodness of the retrieved photos but not the number of retrieved events, nor how accurate the correspondence between retrieved images and events is.
- Normalized Mutual Information (NMI). This compares two sets of photo clusters (where each cluster comprises the images of a single event), jointly considering the goodness of the retrieved photos and their assignment to different events.

Both evaluation measures receive values in the range [0, 1] with higher values indicating a better agreement with the ground truth results.

5. CONCLUSIONS

The SED task gave its participants the opportunity to test and comparatively evaluate different approaches to the problem of social event detection in multimedia collections. The results of the submissions give rise to interesting conclusions. Details on the methods and results of each individual participant can be found in the working notes papers of the MediaEval 2011 Workshop Proceedings.

Acknowledgments

The work presented in this paper was partially supported by the project OpenSEM funded by EIT ICT Labs, and by the European Commission under contracts FP7-216444 Petamedia, FP7-248984 GLOCAL, FP7-215453 WeKnowIt and FP7-249008 CHORUS+.

6. REFERENCES

[1] S. Papadopoulos, C. Zigkolis, Y. Kompatsiaris, and A. Vakali. Cluster-based Landmark and Event Detection on Tagged Photo Collections. *IEEE Multimedia*, 18(1):52–63, February 2011.

[2] A. Scherp, R. Jain, M. Kankanhalli, and V. Mezaris. Modeling, Detecting, and Processing Events in Multimedia. In *18^{th} ACM International Conference on Multimedia*, pages 1739–1740, Firenze, Italy, 2010.

[3] R. Troncy, B. Malocha, and A. Fialho. Linking Events with Media. In *6^{th} International Conference on Semantic Systems (I-SEMANTICS)*, Graz, Austria, 2010.