CERTH @ MediaEval 2011 Social Event Detection Task

Symeon Papadopoulos, Christos Zigkolis
[1]CERTH, Thessaloniki
[2]Aristotle University
{papadop,chzigkol}@iti.gr

Yiannis Kompatsiaris
[1]Informatics & Telematics Institute
CERTH, Thessaloniki, Greece
ikom@iti.gr

Athena Vakali
[2]Informatics Department
Aristotle University, Thessaloniki, Greece
avakali@csd.auth.gr

ABSTRACT
This paper describes the participation of CERTH in the "Social Event Detection Task @ MediaEval 2011", which aims at discovering social events in a large photo collection. The task comprises two challenges: (i) identification of soccer events in the cities of Barcelona and Rome, and (ii) identification of events taking place in two specific venues. We adopt an approach that combines spatial and temporal filters with tag-based location classification models and an efficient photo clustering method. In our best runs, we achieve F-measure and NMI scores of 77.4% and 0.63 respectively for Challenge 1, and 64% and 0.38 for Challenge 2.

Categories and Subject Descriptors
H.3 [**Information Search and Retrieval**]

1. INTRODUCTION
In this paper we present our system, experiments, and conclusions in the context of the MediaEval 2011 Social Event Detection (SED) Task. The SED Task, which is described in detail in [1], provides a collection of 73,645 tagged photos from Flickr and requests the detection of two types of social events. Challenge 1 pertains to the detection of soccer events in the cities of Barcelona and Rome. Challenge 2 asks for events taking place in Paradiso (Amsterdam) and Parc del Forum (Barcelona). The task considers a social event as a group of photos capturing some aspect of a certain event. Formally, given the collection $P \triangleq \{p\}$ of photos, the task asks for the detection of K events $\{E_i | E_i \subset P\}, i = 1, ..., K$.

2. SED APPROACH
We employed a common approach for tackling both challenges. Figure 1 illustrates its main steps: (a) photo filtering, (b) event partitioning, and (c) event expansion.

2.1 Photo filtering
This step is implemented through the cascaded combination of two classifiers. The first classifier is a city-level classifier that is focused on the five cities of the SED dataset. For those photos that have geotagging information associated with them (~20% of the photos [1]), the classifier simply assigns the nearest city to the photo (geodesic distance is used for ranking). For the non-geotagged photos, the classifier employs a tag-based matching scheme to classify the photo to one of the cities: the classifier counts the number of *city-specific* tags in the textual metadata (title, description) of the photo for each dataset city and selects the one, with which the photo shares the maximum number of city-specific tags; the city tags are automatically derived from statistical analysis of tags of city photos collected independently from Flickr. If the classifier assigns a photo to a city that is not of interest for the challenge at hand, then this photo is not further considered (but is not excluded from the event expansion step described in subsection 2.3).

Subsequently, a finer-grained classifier is employed for selecting only the photos that are related to the topic/entity of interest. For Challenge 1, a soccer classifier was created, while a venue classifier was employed for Challenge 2. Both classifiers rely on an approach similar to the one described for the city classification. In both cases, appropriate tag models (soccer model and venue models, one for each of the venues specified by the task) were used that will be further described in Section 3.

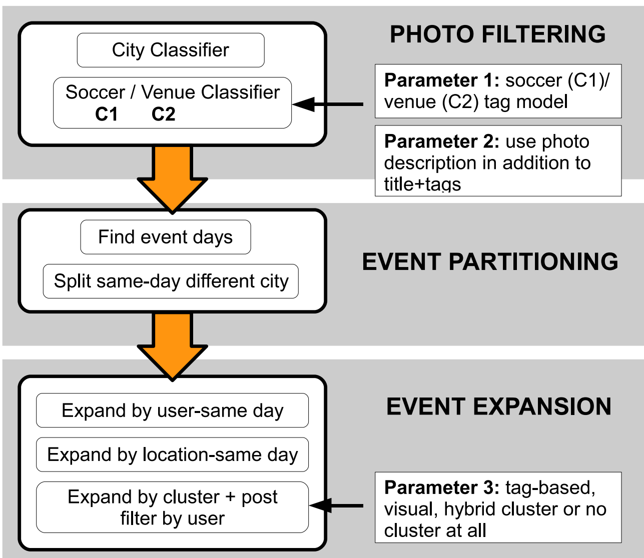

Figure 1: Proposed SED approach

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

2.2 Event partitioning

We define a single event by a date-place combination. For that reason, we first enumerate all unique dates that appear in the set of photos collected from the photo filtering step described above. For each unique date, we consider a distinct event, except for the dates for which there are photos classified to more than one city of interest. For these dates, one distinct event is considered for each different city, with which at least one photo is associated. At the end of this step, a list of events is available and each event of this list is associated with a set of photos.

2.3 Event expansion

Each event produced by the event partitioning step is enriched by making use of the metadata of the photos associated with it. A first expansion is carried out by adding photos of the same user at the same day of the event. Next, photos with geotagging information that are located in the vicinity of the event (within a radius of 200m) are also added to the event under consideration. Finally, an additional list of photos related to the event are discovered by means of clustering the photo collection and selecting the photos of the same cluster under the constraint that their creators/owners are already associated to the event through at least one photo. The photo collection is clustered by means of a community detection scheme that is applied on a visual, tag or hybrid similarity graph [2].

3. EXPERIMENTS

We present a set of 10 experiments that evaluate the performance of the system under a variety of configurations. Tables 1 and 2 summarize the results obtained from the official submission of the five runs to Challenges 1 and 2 respectively. In both tables, the run number along with the selected parameters are listed together with the achieved performance scores, Normalized Mutual Information (NMI), Precision (P), Recall (R) and F-measure (F), which are described in [1]. For Challenge 1, all three parameters appearing in Figure 1 are studied, while for Challenge 2, only the first and third parameter are studied.

The first parameter (p_1) pertains to the tag model used for filtering out irrelevant photos. Two different soccer tag models were used for Challenge 1, i.e. $p_1 \in \{m_{1,b}, m_{1,+}\}$, where $m_{1,b}$ is the baseline soccer tag model containing generic soccer tags as well as tags consisting of spanish and italian football club names. The extended tag soccer model ($m_{1,+}$) additionally contains alternative team names (e.g. "Blaugrana" for Barcelona FC) and stadium names. For Challenge 2, a similar selection was available: the baseline venue tag model ($m_{2,b}$) consisted of few tags with generic music event terms (e.g. "concert", "gig") as well as the names of the two venues of interest. The extended model ($m_{2,+}$) was enriched with the names of the bands playing in these venues in May 2009 that were retrieved by use of the last.fm API.

The second parameter (p_2) regards the use of description along with the photo title/tags ($p_2 \in \{tt, ttd\}$), ttd denoting the use of description in addition to title/tags (tt). The third parameter (p_3) regards the use of clustering for event expansion ($p_3 \in \{\varnothing, T, V, H\}$), where the options of tag-based (T), visual (V), hybrid (H) clusters (produced by graph-based clustering on graphs comprising both tag-based and visual similarities) or no clusters at all ($\varnothing$) were available.

run	p_1	p_2	p_3	NMI	P	R	F
1	$m_{1,b}$	ttd	T	0.3742	57.66	62.50	59.98
2	$m_{1,+}$	tt	∅	0.5707	90.58	**67.58**	**77.40**
3	$m_{1,+}$	tt	T	0.6180	90.58	**67.58**	**77.40**
4	$m_{1,+}$	tt	V	0.5748	89.18	**67.58**	76.89
5	$m_{1,+}$	ttd	T	**0.6301**	**94.63**	65.43	77.37

Table 1: Results for Challenge 1

run	p_1	p_3	NMI	P	R	F
1	$m_{b,2}$	∅	0.2516	51.36	48.85	50.08
2	$m_{b,2}$	T	0.2629	50.58	48.85	49.70
3	$m_{b,2}$	V	0.2527	51.27	48.85	50.03
4	$m_{b,2}$	H	0.2646	50.58	48.85	49.70
5	$m_{+,2}$	H	**0.3796**	**54.31**	**77.90**	**64.00**

Table 2: Results for Challenge 2

4. DISCUSSION

The first important observation by studying the results in Tables 1 and 2 highlights the importance of using an appropriate tag model for photo classification. A significant improvement in all evaluation measures is achieved by use of an enriched tag model. For instance, in Challenge 1, this is clearly visible by comparing runs 1 and 5, while in Challenge 2 it is demonstrated by comparing runs 4 and 5. This highlights the value of rich domain knowledge in the reliable detection of social events in photo collections. In addition, the use of description (*ttd*) in addition to the title and tags of photos appears to improve the performance of our system when the extended soccer tag model is used (compare runs 3 and 5 in Table 1).

Finally, the use of image clustering appears to be of limited value to the system. In Challenge 1, there is a modest improvement in the obtained NMI when using the tag-based clusters (run 3 versus run 2), and a marginal improvement when usign the visual clusters. However, a slight drop in precision is observed when using the visual clusters (run 4 versus run 2). Similar results are obtained for Challenge 2. We attribute this result to the fact that most of the potential gains of clustering are already captured by the user-based event expansion (used in all runs).

In conclusion, the experiments indicate the importance of textual metadata of photos in combination with rich domain knowledge for the effective detection of events in large photo collections. In the future, we plan a more comprehensive analysis of false positives and false negatives in order to further improve the system performance.

Acknowledgments

This work has been supported by the GLOCAL EC project under contract number FP7-248984.

5. REFERENCES

[1] S. Papadopoulos, R. Troncy, V. Mezaris, B. Huet, and I. Kompatsiaris. Social event detection at mediaeval 2011: Challenges, dataset and evaluation. In *MediaEval 2011 Workshop*, Pisa, Italy, September 1-2 2011.

[2] S. Papadopoulos, C. Zigkolis, Y. Kompatsiaris, and A. Vakali. Cluster-based landmark and event detection for tagged photo collections. *Multimedia, IEEE*, 18(1):52 – 63, jan. 2011.