A Generic Approach for Social Event Detection in Large Photo Collections

Matthias Zeppelzauer
Vienna University of Technology, Institute for Software Technology and Interactive Systems
mzz@ims.tuwien.ac.at

Maia Zaharieva
University of Vienna
Research Group Multimedia Information Systems
zaharieva@cs.univie.ac.at

Christian Breiteneder
Vienna University of Technology, Institute for Software Technology and Interactive Systems
cb@ifs.tuwien.ac.at

ABSTRACT

In this paper we explore the performance of a generic methodology for the detection of social events in large photo collections. The proposed approach represents a generic event retrieval system applicable to arbitrary queries and event types. It exploits available textual and visual information for the iterative clustering of relevant spatio-temporal clusters corresponding to event type, location, and time.

1. INTRODUCTION

The continuing increase of available multimedia data and the enormous variety of possible search queries require for generic and efficient indexing and analysis methods that are not tailored to the characteristics of a specific query. The goal of the social event detection (SED) task in the MediaEval benchmark 2012 is to find clusters of photos that belong to the same event in a large Flickr photo collection. A typical query in this context may be: "find photos of all technical events that took place in Germany in March 2012 in the collection". [4] gives a detailed description of the employed photo collection and the challenges (queries) of the SED task. In this paper, we present our approach for social event clustering and report the results obtained in the task. The proposed approach does not make assumptions about the query and, thus, is applicable to arbitrary queries and event types. Except for some general metadata preprocessing, the approach uses only the available textual and visual information.

2. RELATED WORK

In 2011, seven groups participated in the SED task. Six out of seven evaluated approaches build upon specialized processing steps or even separate methods optimized for the actual challenge. For example, Brenner and Izquierdo extract soccer clubs and stadium names from webservices for challenge 1 and music-related venues, events, and artists from Last.fm for challenge 2 [2]. The presented approaches in 2011 are highly specialized for the two challenges from SED 2011 and are not generally applicable to arbitrary queries and event types. While the performance of such approaches is relatively high for the evaluated challenges, they are prone to overfitting and their applicability in a real-world scenario is questionable. Solely one submission in 2011 presents a more general approach for social event detection. Morchid and Linarès use only the information contained in the query and in the provided metadata excerpt from Flickr [3]. The resulting retrieval system is applicable to both challenges without modifications. While the performance is significantly lower than that of the specialized approaches, it gives a more realistic performance estimate of a generally applicable approach.

In our understanding, it is not reasonable to develop specific methods for different queries. In the course of the SED 2012 task we evaluate how well a generic event retrieval approach performs. The goal is not to maximize performance measures but to evaluate the potential of such an approach across different queries and event types. Thus, we develop one generic event clustering approach and apply it to all five challenges in the SED 2012 task[1] using the same parameters.

3. APPROACH

An overview of the proposed approach is given in Figure 1. The core idea is to combine spatio-temporal clustering with filtering and refinement steps. We first cluster the data based on the most reliable information (timestamps and geotags) to obtain robust event candidates. Next, we employ additional contextual information (e.g. user-defined tags, title, visual content) for filtering and refinement of the event candidates.

The two major inputs of the approach are the query and the Flickr metadata excerpt provided by the SED 2012 organizers. From the query we manually extract the query type, location, and time. For test challenge 1 of SED 2012, for example, the type is "technical", the location is "Germany" and time is unspecified.

The provided Flickr metadata is input to a preprocessor which semantically annotates the data (*metadata preprocessing*). We employ GATE (A General Architecture for Text Engineering)[2] to extract information such as dates, locations, person names, and nouns (potential keywords) from the free-text metadata descriptions. Additionally, we process the GeoNames database[3] for the interpretation of geolocation tags. Following, we cluster the photos by their capture date (*temporal clustering*). We perform a simple

[1]SED 2012 task comprises three new test challenges for the 2012 edition and two challenges for the development set which are identical with the test challenges in SED 2011.
[2]http://gate.ac.uk
[3]http://www.geonames.org

MediaEval 2012 Workshop, October 4-5, 2012, Pisa, Italy

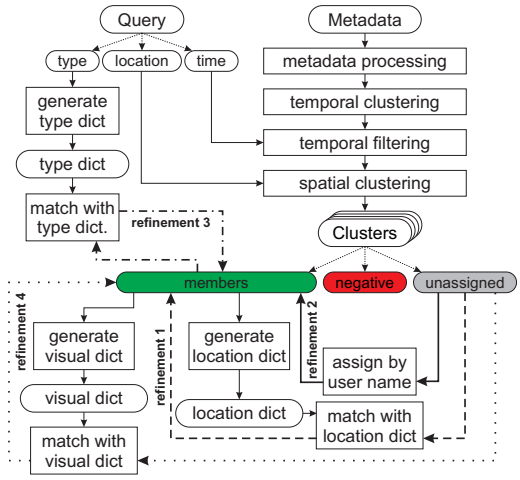

Figure 1: Overview of the approach.

day-by-day segmentation of the images which yields clusters that each represent one day. The major limitation of the approach is that events that last several days are over-segmented. Next, all clusters, which do not coincide with the temporal constraints specified in the query, are filtered out (*temporal filtering*). The remaining clusters are temporally coherent but may represent different locations. We split the clusters into spatio-temporally coherent clusters by assigning the cluster members to one of the locations specified in the query. Cluster members without geo-information are kept in a list of unassigned photos. Cluster members with a different location than specified in the query are removed (*negative list*).

The subsequent process consists of four refinement steps applied to each of the spatio-temporally coherent clusters. The *first refinement step* assigns non-geotagged unassigned photos to the clusters by matching their textual metadata with location-specific dictionaries built from the cluster members. The *second refinement step* adds unassigned photos captured by the same user and the same time as a cluster member to the corresponding cluster. In the *third refinement step* we match cluster members and optionally also unassigned photos (extended type match) with a dictionary that represents keywords related to the query type. Only the best matching photos (with a Jaccard index above a certain threshold t_{type}) remain in the cluster. The *fourth refinement step* exploits visual information to assign photos from the unassigned list to the clusters. We use PHOW features (dense SIFT descriptors)[1] to construct a bag-of-features model from photos that are classified as relevant by the previous steps of the approach. Following, each image from the unassigned list is assigned to a cluster if the chi square distance to its model is below a predefined threshold.

4. EXPERIMENTS AND RESULTS

We perform five runs for all five challenges: two development challenges and three test challenges. For each challenge the corresponding data set from the SED 2012 task is employed. For each run all parameters are identical for all challenges. For runs 1-3 we vary the threshold t_{type} for matching the query type (run 1: 3, run 2: 2, run 3: 4). In run 4 we skip refinement step 2 and perform an extended type match. In contrast to all previous runs, run 5 additionally performs visual matching. Table 1 summarizes the results for all challenges and all runs.

challenge \run		1	2	3	4	5
dev 1	F1	67.87	53.02	78.61	67.87	43.44
	NMI	0.39	0.30	0.48	0.39	0.23
dev 2	F1	47.96	53.32	43.67	47.16	38.21
	NMI	0.21	0.21	0.16	0.19	0.15
test 1	F1	0.61	2.15	0	0.56	0.55
	NMI	0.01	0.02	0	0.01	0.01
test 2	F1	6.46	22.99	5.95	6.38	6.35
	NMI	0.06	0.20	0.06	0.06	0.06
test 3	F1	37.43	47.58	29.33	36.83	36.72
	NMI	0.26	0.31	0.22	0.26	0.25

Table 1: Results for all challenges and five runs (F1 score and normalized mutual information, NMI).

The results achieved on the development set show significant performance improvement of our generic algorithm in comparison to the one presented in SED 2011 (reported F-score for challenge 2 is 3.53 in 2011, and 53.32 using our approach). These first results prove the feasibility of the application of a generic algorithm for the detection of such high-level information as social events.

The main reason for the significant drop in the performance in SED 2012 is the quality of the query keywords. For example, the event type of challenge 1 is *"technical"* which lacks in expressiveness. Challenge 2 introduces ambiguity in the interpretation of keywords (e.g. Barcelona as location vs. Barcelona as soccer club name). Results for challenge 3 are better since the query provides more details (more keywords). Finally, the introduction of visual matching (run 5) has low influence on the performance. This is due to the fact that visual models are based on clusters containing (still) a significant amount of falsely assigned images.

5. CONCLUSIONS

We presented a generic method for the detection of arbitrary social event types. The approach exploits only the provided benchmark and employs no external data, pre-trained models, and challenge-specific processing. Evaluations show that the performance highly depends on the amount of information provided in the query. Future work will explore query expansion and visual model enhancement to improve the overall performance of the approach.

6. REFERENCES

[1] A. Bosch, A. Zisserman, and X. Muñoz. Image classification using random forests and ferns. In *IEEE Int. Conf. on Computer Vision*, pages 1–8, 2007.

[2] M. Brenner and E. Izquierdo. Mediaeval benchmark: Social event detection in collaborative photo collections. In *CEUR Proc. of the MediaEval 2011*, 2011.

[3] M. Morchid and G. Linarès. Mediaeval benchmark: Social event detection using lda and external resources. In *CEUR Proc. of the MediaEval 2011*, 2011.

[4] S. Papadopoulos, E. Schinas, V. Mezaris, R. Troncy, and I. Kompatsiaris. Social Event Detection at MediaEval 2012: Challenges, Dataset and Evaluation. In *MediaEval 2012 Workshop*, Pisa, Italy, 2012.