A Hierarchical Approach to Event Discovery from Single Images using MIL Framework

K. Ahmad, F. G. B. De Natale, G. Boato, and A. Rosani

Multimedia Signal Processing and Understanding Lab, DISI, University of Trento, Italy

IEEE GlobalSIP
Fourth IEEE Global Conference on Signal and Information Processing
December 7-9, 2016, Greater Washington D.C., USA
http://2016.ieeeglobalsip.org/
IEEE Signal Processing Society

Abstract

In this poster, we propose to face the problem of event detection from single images, by exploiting both **background information** often containing revealing contextual clues and details, which are salient for recognizing the event. Such details are **visual objects critical to understand the underlying event** depicted in the image and were recently defined in the literature as "**event-saliency**". Adopting the **Multiple-Instance Learning (MIL)** paradigm we propose a **hierarchical approach** analyzing first the entire picture and then refining the decision on the basis of the event-salient objects. Validation of the proposed method is carried out on two benchmarking datasets and it demonstrates the effectiveness of the proposed hierarchical approach to event discovery from single images.

Introduction

User generated media contents are often associated with personal experiences or social events, by means of which people tends to annotate them for future retrieval. This gives an idea about the big potential of event-centered data analysis. Therefore, in this poster we aim to provide a better solution to event detection in single images

The main contributions of this work are summarized as follows:

(1) We provide a detailed analysis of event-related multimedia contents focusing on key visual elements and background information.

(2) We propose a two step hierarchal approach based on MIL paradigm where initially full images are utilized for event detection in single images, and then event salient features are exploited to further refine the classification.

(3) In the refinement phase, the concept of event saliency, introduced in our previous work Rosani et al. [1], has been utilized to extract event salient objects from event related images.

Proposed Approach

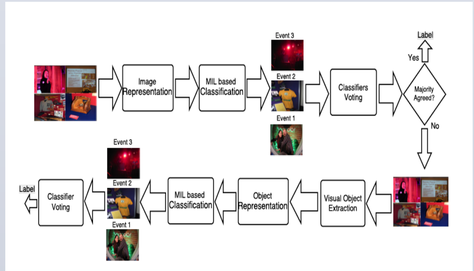


Fig. 1 The flowchart of the proposed framework

Steps of the proposed approach: Our hierarchical approach is mainly composed of 5 steps, which are

1. **Image Representation:** To this aim, Color Structure Descriptor (CSD) is used.
2. **MIL based Classification:** In this work, we adopted C-KNN implementation of MIL paradigm
3. **One-against-one strategy:** To deal with multiclass classification, one-against-one strategy is adopted
4. **Decision Module:** This module decides to switch between the type of information in training samples (i.e., if an image is not correctly classified with full images, event salient objects are used to refine the classification
5. **Extraction of Event-salient Objects:** For this purpose, event saliency maps are used.

Event Saliency

(a) (b) (c)

Fig. 2 Visual Objects Extraction: column (a) represents original images, (b) represent their corresponding event saliency maps while visual objects extracted via event saliency maps are shown in (c)

Datasets for Validation

Datasets: The experimental validation of the the proposed approach has been carried out on three different large scales datasets including EiMM [2], SED 2013 [3], and sub sets of EiMM and SED datasets used in [1].

Event Saliency Maps for 14 classes of Events

SED Data Set	EiMM Data Set
Concert	Concert
Conference	Graduation
Exhibition	Mountain Trip
Fashion	Meeting
Sports	Picnic
Protest	Sea Holiday
Theater	Ski Holiday
-	Wedding

Table 1 List of event classes for which event saliency maps are available

Sample Images and corresponding saliency maps from the Dataset

Fig. 3 Sample images and their corresponding event saliency maps from the dataset

Experimental Settings

Parameters: In our proposed approach, there are two parameters need to be decided

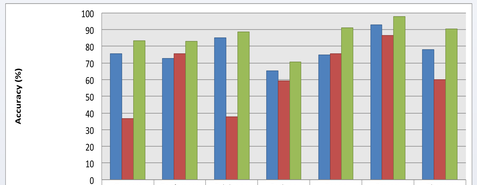
- **Number of images per bag :** In order to analyze the trade off between classification performance and processing time (as processing more images will take longer), we used different configurations on the validation set.
- **Number of Citers and Reference bags:** In order to find best values for these two parameters we tried out different combinations of citers and reference bags on validation set.

Experimental Results

Experimental Configuration: In evaluation process, we conducting experiments with three different configurations which are:

- **Experimental Configuration 1:** In the first configuration, we used a set (5 per bag) of full images (i.e., including background information) per bag in training set while in test set a single image is considered as a bag.
- **Experimental Configuration 2:** In the 2nd experimental configuration, we used the event salient objects, extracted through event saliency maps, from each training sample as the instances of a bag while test bags are composed of full images only.
- **Experimental Configuration 3:** In the third experiment, we used a hierarchical approach where first full images are used for classification purposes, and then event salient objects are used to further refine the classification.

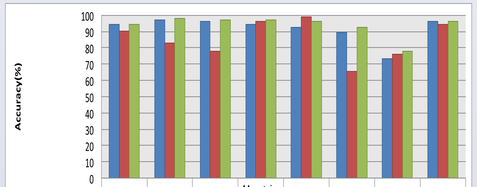
Evaluation on SED Dataset [3]



	Concert	Conference	Exhibition	Fashion	Sports	Protest	Theater
Full Images	75.71	72.86	85.01	65.34	74.91	93	78.13
Visual Objects	36.52	75.47	37.82	59.13	75.61	86.55	60
Hierarchical Approach	83.46	82.94	88.63	70.63	91.08	97.87	90.23

Fig. 4 Experimental results of our approach with all three configurations on SED dataset [3] in terms of accuracy per event class.

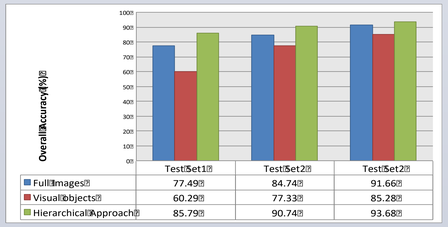
Evaluation on EiMM dataset used in [1]



	Concert	Graduation	Meeting	Mountain-trip	Picnic	Sea Holiday	Ski Holiday	Wedding
Full Images	94.28	97.14	96.19	94.28	92.38	89.52	73.33	96.19
Visual Objects	90.4	82.8	78	96.1	99	65.7	76.1	94.2
Hierarchical Approach	94.28	98.09	97.14	97.14	96.19	92.38	78.09	96.19

Fig. 5 Experimental results of our approach with all three configurations on a dataset used in [1] in terms of accuracy per event class.

Comparison of all of the three experimental configurations in terms of overall accuracy on three different datasets



	Test Set 1	Test Set 2	Test Set 2
Full Images	77.49	84.74	91.66
Visual Objects	60.29	77.33	85.28
Hierarchical Approach	85.79	90.74	93.68

Fig. 6 A comparison of all of our three experimental configurations on three different datasets. Test set 1 represents SED [3] while Test set 2 and Test Set 3 are used in [1]

Comparison of all of our three approaches with a state-of-the-art method [1]

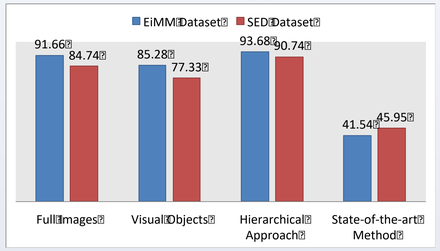


Fig. 7 Comparison with SoA on EiMM and SED datasets used in [1]

Conclusion

In this work, we investigated the importance of background information and event-salient details in the context of event detection. In particular, we proposed a hierarchical approach, incorporating both full images with background information and event-salient details in the Multiple-instance Learning (MIL) framework.

From the experiments, we can conclude:

(1) Though background information can play a vital rule in event detection in single image, sometimes it lead to mis-classification due to visual correlation among certain event classes.

(2) Event salient objects are affective in differentiating among event classes having similar backgrounds.

(3) Using a combination of both full images containing background information, and event salient features leads to a better performance

References

[1] Rosani, A. and Boato, G. and De Natale, F. G B, " EVENT Mask: a game based framework for event saliency identification"IEEE Transactions on Multimedia, 2015.

[2] M. S. Dao, D. T. Dang-Nguyen, and F. G. B. De Natale, "Robust event discovery from photo collections using signature image bases (sibs)," Multimedia Tools and Applications , vol. 70, no. 1, pp. 25–53, 2014.

[3] T. Reuter, S. Papadopoulos, G. Petkos, V. Mezaris, Y. Kompatsiaris, P. Cimiano, C. de Vries, and S. Geva, "Social event detection at mediaeval 2013: Challenges, datasets, and evaluation," in Proceedings of the MediaEval Multimedia Benchmark Workshop , 2013.