J. Vis. Commun. Image R. xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Event recognition in photo albums using probabilistic graphical models and feature relevance ☆

Siham Bacha [a], Mohand Saïd Allili [b,*], Nadjia Benblidia [a]

[a] LRDSI Laboratory, Saad Dahlab University - Blida1, Blida, Algeria
[b] Department of Computer Science and Engineering, University of Quebec in Outaouais, Gatineau J8X 3X7, QC, Canada

ARTICLE INFO

Article history:
Received 9 March 2016
Revised 25 June 2016
Accepted 24 July 2016
Available online xxxx

Keywords:
Photo albums
Event recognition
Object/scene relevance
Probabilistic graphical models (PGM)

ABSTRACT

This paper proposes a method for event recognition in photo albums which aims at predicting the event categories of groups of photos. We propose a probabilistic graphical model (PGM) for event prediction based on high-level visual features consisting of objects and scenes, which are extracted directly from images. For better discrimination between different event categories, we develop a scheme to integrate feature relevance in our model which yields a more powerful inference when album images exhibit a large number of objects and scenes. It allows also to mitigate the influence of non-informative images usually contained in the albums. The performance of the proposed method is validated using extensive experiments on the recently-proposed PEC dataset containing over 61 000 images. Our method obtained the highest accuracy which outperforms previous work.

1. Introduction

The proliferation of digital cameras has contributed to produce an increasing amount of personal photos with an exponential rate. Therefore, the need for efficient and advanced methodologies regarding personal photo collections management emerges as a challenging and imperative issue. In the last decades, a number of research works have focused on the development of techniques for effective organization of personal photo collections [3]. These works process image visual content to infer high-level semantics as perceived by humans [41]. Researchers have incorporated semantic cues, such as faces [5,32] and person identification [33] to help photo collections management. Moreover, contextual cues such as time-stamps and GPS information have been also used for the same objective [15,21,31,46,49].

In real-world scenarios, people usually take photos that are related to particular events (e.g., *birthdays*, *sport events*, etc.), and the photos are arranged later on into albums. Events can also be considered as an important semantic clue for recalling photos content [60]. Therefore, automatic event recognition in personal photo collections plays an important role for intelligent photo management and advanced retrieval. It is also important for applications such as semantic image indexing and summarization [15,18] and security enforcement [33]. Several methods have been proposed to deal with event recognition on single images or group of photos. For example, methods based on bag-of-features models have been used to predict event categories [10,17]. Recently, features based on deep Convolutional Neural Networks (CNNs) [22]) have been successfully used with classifiers such as neural networks [52] and nearest neighbor [38] for event prediction. In addition to visual features, contextual information (e.g., time-stamps, GPS, etc.) has been also used to enhance event recognition [8,23,39,45,55]. The major limitation of the above methods, however, is that they heavily rely on classifiers based on low-level features. Since these features have no explicit semantic meaning and can be shared by several events, event recognition becomes less efficient and interpretable.

To address this issue, several works propose to use high-level semantically meaningful features for event recognition. To recognize events, [6] use correlation between scene categories (e.g., mountains, urban areas, etc.) and events. Although scene information can provide some good clues about events, it is insufficient to discriminate events sharing the same scene categories. For example, *Wedding* and a *Birthday* events can be associated to the same scene types, but can contain different objects. Therefore, foreground objects are important for event recognition. For instance, a *Hiking* event can be intuitively derived from a 'snowy mountain' scene, whereas a *Wedding* event is usually characterized by the presence of white-dressed 'bride'. To integrate object information, another line of works has been proposed recently for event

☆ This paper has been recommended for acceptance by Zicheng Liu.

* Corresponding author.
E-mail address: mohandsaid.allili@uqo.ca (M.S. Allili).

http://dx.doi.org/10.1016/j.jvcir.2016.07.021
1047-3203/

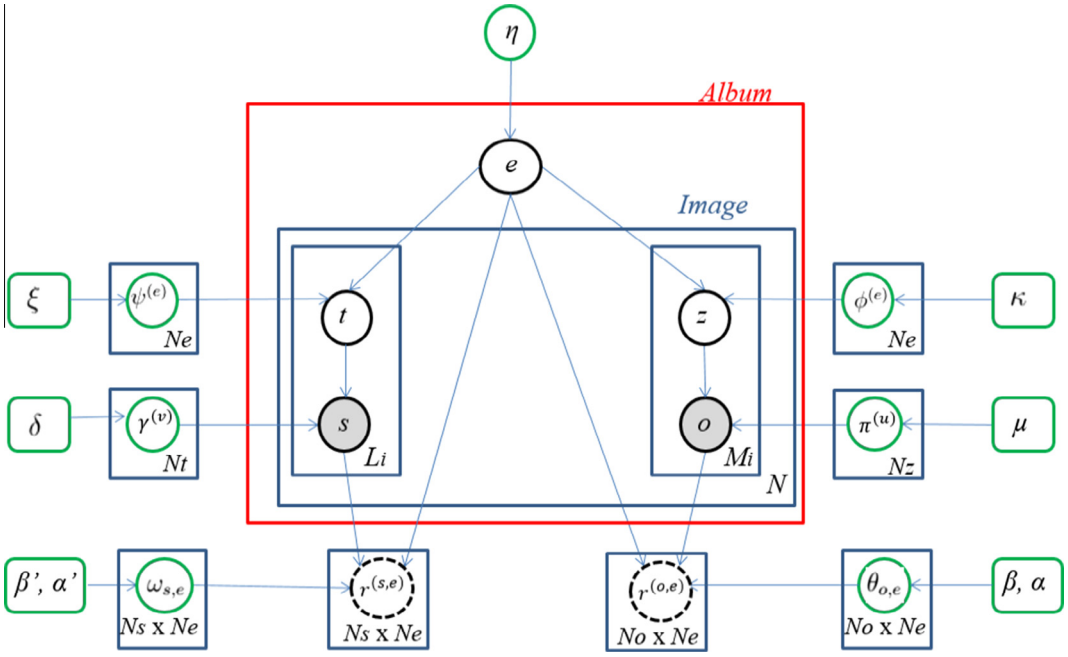

Fig. 1. Graphical representation of our model. Boxes denote replication of the corresponding random variables: there are N images in an album $\mathcal{A}$, with M_i observed objects and L_i observed scenes in image I_i. The variables e, o and s represent the event, object and scene classes, respectively. The latent variables for generating object and scene are z and t, respectively.

recognition. For example, [47] propose a statistical method selecting representative objects for event categories. For each object category, a detector is built and its response used to predict event categories using SVM. Another work from the same authors [48] propose to mine frequent object pattern to determine the most discriminative ones. A photo album is then expressed as frequencies of this discriminative pattern, which they called Compositional Object Pattern Frequency. However, since this work does not use scene information, classes containing similar objects but different scenes can be confused. For instance, *City-tour* and *Motor-show* events can contain the 'car' object but the 'car' appears in different contexts. In other words, it is difficult to generate accurate event semantics based only on object representation. Contextual information is often required to understand the role and dynamics of objects in the events [26].

Although several approaches have been proposed to use either scene categorization or object detection for event recognition, the possibility of jointly using these two tasks to recognize events has not been well investigated. In [26], a statistical model integrating scene and object information for event recognition on single images is proposed. Recently, [50] propose using features derived from CNNs [22] to perform event recognition in groups of images. In the same vein, [28] employ CNNs features to extract scene and object information. Then, event categories in images are predicted using discriminant analysis. It remains, however, that these methods are more adapted on detecting events on single images which provide less rich and complete information about events compared to entire albums. On the one hand, photos constituting albums are usually taken at times-tamps reflecting important moments of the events [3] and, therefore, can be more informative about the events. On the other hand, different photos can give a more exhaustive set of objects/scenes involved in the events.

In this paper, we propose a combined object/scene-based approach for event recognition in personal photo albums. More specifically, a probabilistic graphical model (PGM) is proposed to infer event categories by leveraging high-level information about objects and scenes in sets of images. Given an album of images, we use CNNs first to extract object and scene information in each image of the album which constitute our high-level features for event recognition. We propose a PGM to model relationships between events and object/scene categories. In the same model, we propose to integrate object/scene relevance to boost event recognition. Finally, to infer event categories of new albums, we combine features of their images in a penalized-likelihood function and use the maximum a posteriori probability (MAP) to estimate their event categories.

Contrary to recent work for event recognition in albums [3,8], our approach relies only on visual information, although using time or localization can be included in later versions of this work to improve performance. Moreover, we combine object/scene cues and incorporate their discriminative power (i.e., relevance) for more efficient event recognition. For instance, a 'stage' object may help to determine the event *Concert*, but not the event *skiing*. Similarly, a 'Christmas tree' can tell much about a the event *Christmas* than other objects. Therefore, including object/scene relevance can be a very useful tool to yield more accurate and interpretable event recognition. Our key contributions in this paper can be summarized as follows:

(1) We propose a probabilistic graphical model (PGM) for event recognition in photo albums by combining scene and object cues. Since Rand [35], several works have discussed how event perception occurs in the human vision system (HVS) [4,13,29]. The advent of Bayesian methods has offered a new an attractive tool to explore the potential of modeling visual inference in a closer way to the HVS. Therefore, using PGMs is a promising tool to explore for event recognition on sets of images.
(2) We introduce a feature relevance (FR) scheme incorporating the predictive power of each object/scene for event recognition. Using FR is motivated by several findings of psychological studies in human cognition, where real-world scene understanding at large extent relies on analyzing objects and their contexts [7,12,43,56]. Most often, people can

Please cite this article in press as: S. Bacha et al., Event recognition in photo albums using probabilistic graphical models and feature relevance, J. Vis. Commun. (2016), http://dx.doi.org/10.1016/j.jvcir.2016.07.021