Neurocomputing 172 (2016) 159–167

Contents lists available at ScienceDirect

Neurocomputing

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

Segmentation and recognition of multi-model photo event

Feibin Yang [a], Qinghua Huang [a,*], Lianwen Jin [a], Alan Wee-Chung Liew [b]

[a] School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510641, PR China
[b] School of Information and Communication Technology, Gold Coast Campus, Griffith University, Nathan 4222, QLD, Australia

ARTICLE INFO

Article history:
Received 18 November 2013
Received in revised form
1 May 2014
Accepted 6 August 2014
Communicated by "Chennai Guest Editor"
Available online 12 May 2015

Keywords:
Multi-model
Event segmentation
EM learning
Picture collage

ABSTRACT

Photograph becomes more and more convenient with the rapid growing of smart phones, which makes photo become one of the most widely used social media. In this paper, we have proposed a segmentation and recognition method for multi-model photo event to help users organize and manage their increasing photo collections. Generative model of photo event is built by analyzing time, location, camera parameters and visual content of photos. Expectation-Maximization (EM) learning algorithm is applied to discover the best parameters of the proposed generative model. With the defined model, each photo is categorized into the corresponding event by calculating the maximum posteriori probability. The representativeness of an event is a photo collage constructed by selecting a set of representative photos from the corresponding event. The proposed method is characterized by the following properties: (1) unlike most of the existing photo event segmentation methods, the location of photos is treated as a key feature, (2) the representativeness of an event is a picture collage instead of a single photo, which is not only informative but also very appealing. The experimental results show that the proposed method is effective and efficient on the photo collections from three experienced smart phone users.

1. Introduction

With the rapid development of Internet and the smart phone, more and more people like to take photos to record their life and share with their friends. Besides, the appearance of the social media sharing services such as Flickr [1] and Pinterest provides people unlimited storage space to collect their photos. Due to these conveniences, the photo collections taken and shared in the social media network is growing so fast that the organization and searching of photos [2,3] become a quite difficult and fussy job. It is natural to segment all the photos in a collection into different categories and sort them according to a predefine criterion. In another words, it is a common need to segment the photos into episodes or meaningful events, in which an event is defined as a group of photos with not only strong similarity in image content but also relatively close proximity in time and location.

There is a large body of work focusing on the photo event segmentation. To our knowledge, most of the studies in the literatures partition the photo collection into different events based on timestamp of photos. A simple method is to take the time gap as the segmented boundary when the time gap is greater than a predefined threshold. Platt [4] proposed a time-based clustering algorithm that segments a photo collection into different events using a threshold of one hour. Later, the author adopted an adaptive threshold scheme [5]. Loui et al. [6] proposed another time-based clustering algorithm. The created time of photos were extracted as the feature, then a two class *K*-means method was used to segment the photo collection. Graham et al. [7] proposed a hierarchical time-based event segmentation algorithm. The approach first sorts the list of photos with respect to the time and creates initial clusters using threshold strategy. Then the initial clusters are split into finer clusters based on the time differences between photos within each initial cluster. Finally, it merges the finer clusters into more general clusters in terms of the time differences between the initial clusters. Gargi [8] proposed an algorithm that takes a long interval with no photo taken as the end of an event and uses a sharply increasing frequency of photo taking as the start of a new event. These time-based segmentation methods are relatively simple, but often generate wrong clusters due to improper or missing time stamps.

Because of the limitations of using only time-based information, the event segmentation methods using content-based features have also been proposed. The content-based features are able to provide additional information about the photos. Most of the content-based clustering methods partition the photo collection into different categories according to the content similarity. Loui et al. [6] first segmented the photo collection into events using timestamps, and broke each event into sub-events based on similarity of color histogram statistic. Platt [4] also grouped photos into events based on the creation time. If the time-based clusters were too large, the content-based clustering would be applied to

* Corresponding author.
E-mail address: qhhuang@scut.edu.cn (Q. Huang).

http://dx.doi.org/10.1016/j.neucom.2014.08.104
0925-2312/

split each large cluster. Cooper et al. [9] proposed a content-based algorithm which considered the similarity of discrete cosine transform coefficients. All these algorithms take into account the content feature and time feature; however, time and content information are treated separately in these studies. In most of them, the content-based clustering is only used when the time-based clustering cannot achieve desired results.

Several methods have utilized Exchangeable Image File Format (EXIF) metadata. The EXIF metadata is the contextual information recorded with a photo. The studies in [10–12] have emphasized the importance of the EXIF metadata. Gong et al. [10] presented a photo event segmentation method based on the scene brightness which was derived from the EXIF data. Gozali et al. [11] applied a hidden Markov model to segment the photo collection by learning the parameters from the time, the EXIF metadata and visual information. Mei et al. [12] presented a segmentation approach which incorporated the time, the EXIF metadata and visual information into a unified probabilistic framework. Their experiments showed that these segmentation methods which fused the information of time, EXIF, and content to generate different events could achieve effective results.

However, almost all the methods described above mainly focus on the event segmentation of the home photo and few methods utilize Global Positioning System (GPS) information to segment the photo collection. As is well known, location is a very important feature [21] in social network. When people take a photo and share it with their friends, they would like to let their friends know where they are. Besides, location can provide more information to photo event segmentation. As Gong et al. [10] described, location was a very important factor when defining an event. But they did not use it as a clustering feature because of the difficulty to obtain location information.

Fortunately, with very fast and significant progress in development of information technology, GPS has become more and more available. Now most of the mobile phones have the ability to obtain their locations via the build-in GPS sensors. The GPS information can be obtained and recorded together with the images when taking photos. Obviously, the change of location results in the change of event. Motivated by this point, we propose a novel segmentation and recognition method for multi-model photo events. Multi-model means that our method is different from most of the previous works. Unlike these traditional photo event segmentation methods which only consider a single feature, all the features of time, GPS, camera parameters, color and visual content are used to describe a photo in the proposed method. Each feature is assumed to obey an unknown distribution. We group the photos into different categories by fusing the time, GPS, EXIF metadata and visual information distributions into a unified framework.

After segmentation of the photo collection, users would often like to select a representative photo for each event so that they can easily recall the content of the corresponding event at a glance. Obviously, the quality of the selected photo is pretty pivotal for the organization and browsing of photo collection. Cooper et al. [9] simply adopted the earliest photo in an event as the representative photo without considering the content information. Mei et al. [12] selected the photo with maximum *a priori* probability in an event as the representative photo in their unified probabilistic framework. Chu et al. [13] proposed a different representative photo selection method. For a given photo event, they first detected the near-duplicate photo pairs, then modeled the relationships by a graph, finally selected the most typical one by examining the mutual relation. It is observed that most of these methods mainly focus on selecting one representative photo for an event. However, there are often several events each of which contains so large number of photos that a single photo is not adequately representative. In order to tackle this issue, we propose a novel method to present an event with a collage which is constructed using the most representative photos in the event.

In order to address the challenges mentioned above, we propose a novel multi-model photo event segmentation and recognition method. All the features of time, GPS, camera parameters, color and visual content are used to describe a photo in the proposed method. By analyzing the latent relationships of these features and the unobservable events, a generative model of photo event is adopted for photo clustering. We choose the EM learning algorithm to estimate the parameters of the generative model, according to which the photo collection is segmented by categorizing the photos into different events. For each event, the Affinity Propagation (AP) clustering [17] is applied to segment the event into a number of sub-events in terms of the content similarity. Finally, we select the most representative photo from each sub-event to construct a picture collage as the representativeness of the event.

In summary, the main contributions of this work are concluded as follows.

(1) The location of a photo is treated as a key feature when segmenting the photo collection. Different locations of two photos mean that they belong to different events.
(2) A picture collage is regarded as the representativeness of an event instead of a single photo. The collage not only provides more information of the event but also shows the content in a visual appealing way.

The rest of this paper is organized as follows. Section 2 describes the details of the proposed method. Section 3 describes the experiments and evaluation results. In Section 4, a conclusion of this work is drawn.

2. Methods

2.1. Feature extraction

As described above, our algorithm relies on the fusion of time, location, camera parameters and visual information. Most of the information can be obtained from the EXIF. EXIF includes hundreds of metadata, i.e. the date and time when the photo is taken, the GPS location where the photo is taken and other camera parameters (e.g. exposure time, flash, focal length, subject distance, etc.). It means that the EXIF can provide a large amount of information to event segmentation. After analyzing a large photo collection, we mainly focus on the following features.

Time (T): Time is obtained from timestamp of the EXIF. If the timestamp is not available, we take the creation time of the file instead.

GPS (G): GPS contains the location information where a photo is taken. It can also be obtained from the EXIF which is given by longitude and latitude. If the GPS information is not recorded in the EXIF, we assume that it is the same as that of the photo which has the most similar time.

Camera parameters (CP): We only utilize three event-related parameters as reported in [12]. The parameters include aperture, exposure time and focal length. Aperture is the diameter of lens opening when capturing photos. Exposure time is the time for which the shutter is held open to allow light to reach the image sensor. The focal length is the distance from the optical center of the lens to the local point.

In addition to the features extracted from EXIF, visual information which provides abundant description of a photo is also very important to event segmentation. We extract the three most