Image and Vision Computing 32 (2014) 751–760

Contents lists available at ScienceDirect

Image and Vision Computing

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

Exploring album structure for face recognition in online social networks☆

Jason Hochreiter *, Zhongkai Han, Syed Zain Masood, Spencer Fonte, Marshall Tappen

University of Central Florida, 4000 Central Florida Blvd., Orlando, FL 32816, United States

ARTICLE INFO

Article history:
Received 31 May 2013
Received in revised form 25 November 2013
Accepted 9 January 2014
Available online 19 January 2014

Keywords:
Face recognition
Online social networks
Structural SVM

ABSTRACT

In this paper, we propose an album-oriented face-recognition model that exploits the album structure for face recognition in online social networks. Albums, usually associated with pictures of a small group of people at a certain event or occasion, provide vital information that can be used to effectively reduce the possible list of candidate labels. We show how this intuition can be formalized into a model that expresses a prior on how albums tend to have many pictures of a small number of people. We also show how it can be extended to include other information available in a social network. Using two real-world datasets independently drawn from Facebook, we show that this model is broadly applicable and can significantly improve recognition rates.

1. Introduction

Traditional face recognition systems have relied on image features to identify the individuals in photographs. The advent of popular social networks, such as Facebook, which host photo albums and make it possible to tag photos with user identities, adds a new dimension of data that can be used to help identify faces.

In Facebook, as in most photo management services, photos are grouped into albums. These albums are a rich source of information because they often correspond to trips, events, or specific groups of people. In this paper, we show how the organizational structure of photos into albums can be used to significantly increase recognition accuracy. In addition, as will be shown in Section 3, our model based on using album information can be applied to significantly more pictures than other models that exploit co-occurrence in photos, such as [1].

Much of the contribution of this work lies in modeling how individuals tend to co-occur in photo albums. The basic model is constructed with the underlying idea that albums tend to contain multiple photos of a small number of people, such as an album containing photographs from a trip. An album may contain many photos, but it is likely that the individuals pictured in the album will be dominated by the small number of people that participated in the event.

After introducing this basic model, we will then show how it can be improved and extended by considering other factors such as previous co-occurrence in an album, friendship information, and the identity of the person who uploaded the photo to the social network.

The rest of the paper follows this rough outline:

- Section 2 illustrates the difficulties of working with data from publicly available social networks.
- Section 3 shows that a CRF model based on limiting the number of individuals appearing in an album is useful for a significant portion of photos on Facebook. This section will show how this model can also be applied more widely than just modeling co-occurrence in photographs. Following this, Section 4 discusses related work.
- Sections 5 through 7 describe the design of the model, inference with the model, and the training procedure for the model.
- Section 8 describes experiments showing how this model significantly increases recognition performance. Most importantly, for the two datasets we test, this approach increases accuracy by around 20% over a baseline classifier.
- In Section 9, we describe a simple stochastic coordinate descent approach to learn the model parameters. That a technique having such low complexity requirements can achieve comparable results illustrates the flexibility of our model.

2. Reproducible research on social network data

As will be discussed in Section 8, we validate our methods on real data gathered from Facebook. One of the difficulties in working with social network data is the ability to share that data. Common datasets, such as the PASCAL challenge [2] and the Middlebury Stereo database [3] have facilitated significant advances in vision technology. However, sharing social network data is problematic due to privacy issues since it involves sharing the information of both users and their friends on social networks.

☆ This paper has been recommended for acceptance by Lijun Yin, Ph.D.

* Corresponding author.
E-mail address: jasonhochreiter@gmail.com (J. Hochreiter).

0262-8856/$ – see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.imavis.2014.01.002

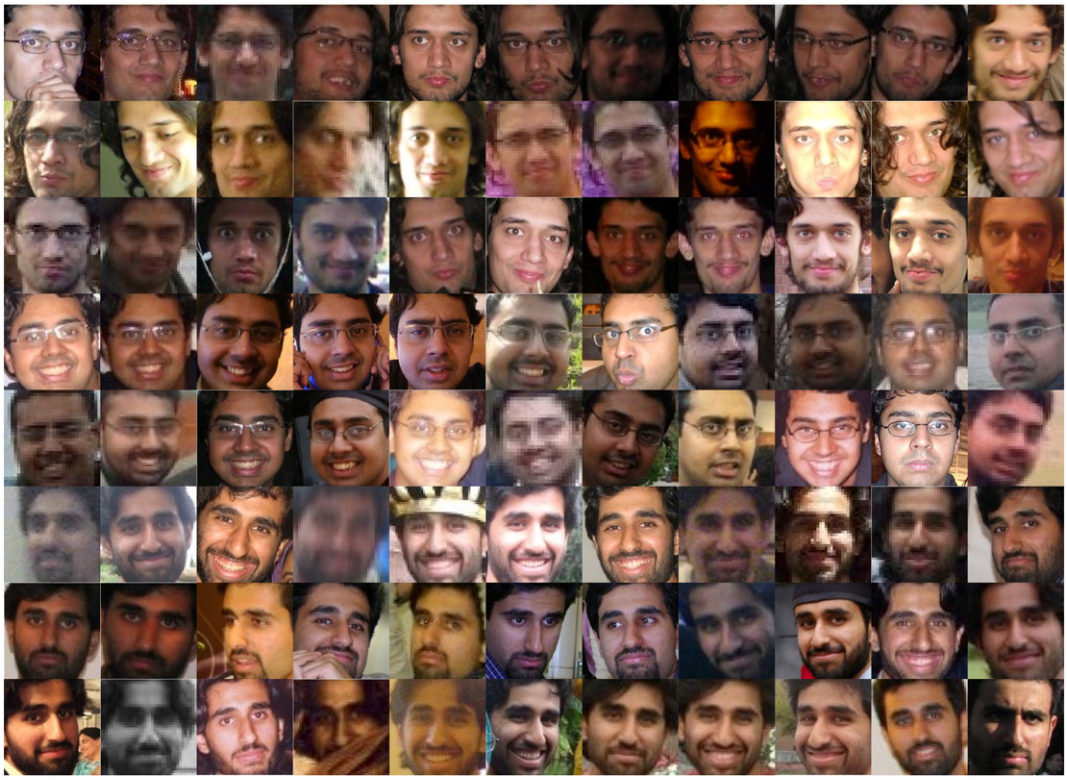

Fig. 1. Sample images for three users from our dataset. The dataset contains a variety of facial poses, expressions, and image quality.

While it could be argued that anonymizing the type of data used for face recognition by various transformations could protect the identities of users, researchers have been effective at de-anonymizing data. In 2008, researchers de-anonymized significant portions of a dataset released by Harvard [4]. Given the constant threat of having anonymized data cracked, future efforts are required to gather realistic social network data that can be distributed.

To more strongly validate our results, we used two different datasets downloaded from the accounts of different individuals and two implementations of our algorithms created by different members of our group. As will be reported in Section 8, both experiments confirm the benefits of the proposed approach.

An example of facial images obtained from three users appears in Fig. 1. These are raw images obtained from Facebook; there is no guarantee of quality or consistency of facial pose and expression in the images.

3. The applicability of an album prior

A key question facing this work is whether users tend to organize albums in a way that makes this prior useful. Our study on photo albums in Facebook indicates that a prior based on the assumption that albums tend to contain multiple photos of a small number of people is much more applicable than a model that relies on co-occurrence inside a photo, such as [1]. To evaluate the usefulness of this prior on the occurrence of individuals in albums, we used the Facebook API to download all pictures visible to a single user's account, similar to [1]. In total, we were able to capture 8078 pictures containing 2849 different people across 1649 albums. In all of the pictures, we only considered faces that had been tagged by a user. To ensure the accuracy of tags, we applied the OpenCV face detector to each tagged photograph we downloaded [5]; if the detector did not indicate the presence of at least one face in a photo, we discarded it. In total, we collected 11,724 facial images.[1]

[1] This data eventually became the first of our two datasets. See Section 8.1.

In this collection of photographs, over 5735 photographs, or 71% of the photographs, only contained one tagged face. Presuming that, for the vast majority of photographs, all of the faces have been tagged, this means that a model based on co-occurrence inside a photograph [1] would help with recognition in roughly only 29% of photographs.

In measuring the applicability of our model, we set a high standard for its usefulness by assuming that an album prior would only be useful if there are at least twice as many photographs in an album as people occurring in that album. Despite this high threshold, we found that an album prior could still be applied to albums containing 57% of the photos in our dataset. Roughly speaking, a prior based on limiting occurrence in albums can be applied to nearly twice as many photographs as a prior just based on co-occurrence inside a photograph, such as the model in [1].

In Section 8.4, we consider the scenario in which this high standard is not met. Training and testing on all albums – for some of these albums, the number of labels is more than half of the number of photographs – results, in only a 2% performance degradation as compared to training and testing on only albums that satisfy this threshold, indicating the applicability of our approach.

Fig. 2 demonstrates why an album-based model can be applied more widely. Each row corresponds to photographs taken from a specific album. For each row, only one photograph has two people in it. Thus, a model based on photograph co-occurrence could only be applied to one of the photographs, while all three could be considered in an album-based model.

An important contribution of this work is that we show a model where the album structure can be considered in an efficient inference scheme.

3.1. Quality of downloaded data

The metadata provided by downloaded photographs and albums may be incomplete or inaccurate. The former case is often due to privacy concerns; the data we are able to download may be limited by the privacy settings of particular users on Facebook. We find this to be of little consequence: even without the capability to download all possible