Computer Vision and Image Understanding 139 (2015) 21–28

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

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

Location recognition over large time lags ☆

Basura Fernando*, Tatiana Tommasi, Tinne Tuytelaars

KU Leuven ESAT-PSI, iMinds Kasteelpark Arenberg 10 – bus 2441 Heverlee B-3001, Belgium

ARTICLE INFO

Article history:
Received 26 September 2014
Accepted 25 May 2015
Available online 25 June 2015

Keywords:
Location recognition
Cross-domain image retrieval
Domain adaptation

ABSTRACT

Would it be possible to automatically associate ancient pictures to modern ones and create fancy cultural heritage city maps? We introduce here the task of recognizing the location depicted in an old photo given modern annotated images collected from the Internet. We present an extensive analysis on different features, looking for the most discriminative and most robust to the image variability induced by large time lags. Moreover, we show that the described task benefits from domain adaptation.

1. Introduction

A hundred year old photograph or a postcard can reveal a lot about our culture and history. Following this idea, many cultural heritage campaigns recently started to promote the digitization of large amounts of visual data. Several cities and towns all over the world, as well as institutions such as universities or museums, are bringing archives with their images and footage online, providing public access and calling for methods to efficiently open up and exploit these resources [1,2].

At the time when photography was not affordable for private and everyday use, most of the pictures were taken in public places and depict buildings, monuments, statues, or more in general, common locations of interest. Some of those are landmarks and tourist attractions. Others are locations with historical value. Popular landmarks often appear in modern digital images which are shared online through applications such as Flickr. Other historical locations can be associated to their geographic coordinates through Google Maps and visualized by means of applications like Google Street-View. Despite the place correspondence, the visual appearance of old and new images is dramatically different. As shown in Fig. 1, ancient photographs have different colors, texture, and contrast characteristics compared to modern digital images [3]. Moreover it is not possible to control the acquisition perspective: changes in the urban planning along the years may have made some viewpoints not accessible.

Numerous efforts have been dedicated to recognizing landmarks in image databases containing photographs of the same era [4–7], but to our knowledge, no previous work focused on tackling location recognition over large time lags. Here we define this task: **annotate an ancient photograph with the correct location label, given a set of labeled modern photos**. In particular, we propose several useful tools to cope with this problem, making three main contributions:

- we introduce a collection of images spanning over 25 locations and more than one century, with the eldest photographs dating back to the 1850s;
- we present a detailed analysis of existing feature representations, looking for the most robust features, suitable to handle the variability induced by different imaging processes adopted over time;
- old and new images can be considered as belonging to two different domains. We use existing domain adaptation methods and we show promising results in both location recognition and interactive location retrieval.

The rest of the paper is organized as follows. Section 2 revises the related work on location recognition and domain adaptation. Section 3 introduces our large time lags locations dataset and indicates the challenges of location recognition on this testbed. Section 4 briefly reviews the domain adaptation methods used in our study. In section 5 we present and discuss the obtained experimental results. Finally, section 6 concludes the paper and points out possible directions for future research.

2. Related work

Location recognition consists in determining where a photo was taken by using as reference a database of previously seen locations [4]. The interest towards this task grew together with the number of freely available images on the Internet, many of which are geotagged and depict urban outdoor scenes. Today, with the widespread use of mobile devices endowed with built-in cameras and Internet connectivity, location recognition is a useful tool for city guides and

☆ This paper has been recommended for acceptance by Frederic Jurie.
* Corresponding author. Tel.: +32163 72422.
E-mail address: basura.fernando@esat.kuleuven.be, basuraf@gmail.com (B. Fernando).

http://dx.doi.org/10.1016/j.cviu.2015.05.016
1077-3142/

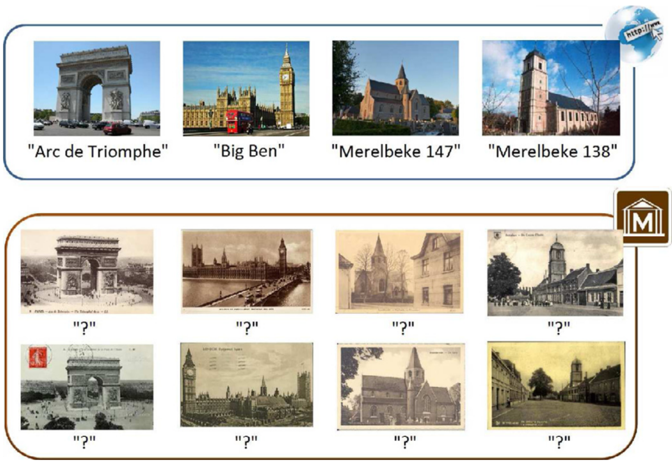

Fig. 1. Pictures of four locations over large time lags showing an evident change in visual appearance. The photographs are similar in their high level scene content, but the color range and texture are significantly different. Modern photos can be easily found on the World Wide Web, while ancient pictures are provided by cultural heritage museums. The task we address in this paper consists in annotating ancient pictures given a set of labeled modern images.

smart navigation aids that are able to localize an image in near real time [8,9].

Given a structured database covering a pre-defined set of places, location recognition can be tackled as a classification problem [5,6]. The models for each place are learned offline and, at query time, a photograph is localized by assigning to it the label of the best scoring location classifier [5]. Previous work also considered this task as a retrieval problem: a query image is used to find a set of similar images from a database which are then returned as place suggestions [7,10,11]. This setting is mainly adopted when dealing with reference image collections possibly containing a large number of distractors.

Regardless of the chosen setup, one of the main challenges for location recognition is the choice of appropriate image descriptors. The variability in illumination conditions, viewpoint and occlusion can dramatically influence the similarity of images even depicting the same place or building. The data similarity is generally based on local descriptors and bag-of-words (BOW) based techniques [12], and the retrieval is performed by computing distances between sparse BOW histograms [13]. Several improvements on this core system have been proposed by learning better descriptors [14,15], introducing more accurate descriptor matching [16], exploiting 3D point clouds as powerful representations [4,17], or carefully handling repetitive structures such as building facades [7].

The mentioned large visual variability occurs in spite of the standard practice of using photos acquired with high resolution modern cameras for location recognition. Although urban scenes and landmarks have been often captured even in ancient pictures and paintings, these samples are generally neglected and the further issues induced by vintage color processes or artistic Brushstrokes are not considered in this task in the literature. One attempt to define robust detectors and descriptors was presented in [18,19], where local symmetry features and spectral correspondence methods are proposed to match urban scenes with lighting, age and rendering style variations. The problems of alignment between paintings and photographs [20,21] and viewpoint re-capturing over time [22] have been tackled mainly leveraging over 3D models. The pioneering work of Shrivastava et al. [23] defined visual similarities between paintings and pictures taken in different seasons. The proposed method relies on the robustness of HOG features [24] and leverages the visual uniqueness of query images against millions of negative data. Despite their relevance, all these approaches have not been tested before for location recognition.

Solving the problem induced by data variability is also one of the goals of *domain adaptation* [25]. Instead of focusing directly on image-pairs matching, domain adaptation examines the data distributions from which the images are drawn. Specifically, two sets of data are considered as belonging to two different domains if they cover the same set of classes but their marginal distributions differ. The aim of domain adaptation is to reduce this distribution shift [25]. Various approaches fulfill this purpose by sample re-weighting and selection [26,27], self-labeling [28,29] and metric learning [30,31]. A solution that has recently received a lot of attention in the computer vision community consists in embedding the samples in a low dimensional subspace shared by both the domains and invariant to their specific characteristics [32–34,46]. This strategy allows to tackle cases where the samples present originally high dimensional feature vectors and one of the two domains contains only unlabeled samples (unsupervised domain adaptation).

Previous work demonstrated that time can naturally cause a visual domain shift [35,36]. Existing methods applied to close this time gap proposed to discover object-specific style-sensitive patches [37], to predict the behavior of time-varying probability distributions [38] or to learn models adaptively over a continuous manifold [36]. However, all these approaches require details about the time ordering (evolution) of images, which is often difficult to obtain, especially with ancient photographs. In many cases only two set of data are available, one older than the other without any further information. Our work fits in this context. We focus on the problem of location recognition over large time lags where we are given a set of labeled modern photos and we want to annotate unlabeled historical pictures.

3. The large time lags locations dataset

As detailed earlier, location recognition has so far been studied over modern images and the issues induced by large time lags have been only marginally considered for other tasks. Therefore one of the contributions of this paper is a database of images which spans over a wide time period and numerous locations. The dataset is presented in this section and used throughout the paper.

3.1. Details of the dataset

We introduce here our large time lags locations (LTLL) dataset containing pictures of 25 locations captured over a range of more than 150 years. Specifically, we collected images from several cities and towns in Europe such as Paris, London, Merelbeke, Leuven and ancient cities from Asia such as Agra in India, Colombo and Kandy from Sri Lanka. We chose thirteen locations considering the presence of well known landmarks for which it has been easy to download old and new pictures from the Web. The remaining twelve locations are in the municipality of Merelbeke, Flemish Province of East Flanders in Belgium. Ancient images of these historical locations dating back to the period 1850s–1950s have been provided by the city archive of Merelbeke. We downloaded the corresponding modern images from Flickr, Google Street-View and the Google-Images search engine, although for some of the locations only a limited amount of modern photos could be obtained. Some statistics about the dataset is shown in Table 1.

Table 1
Some dataset statistics. Minimum, maximum and mean number of images per class is shown.

Image set	Minimum	Maximum	Mean
New images	4	22	11
Old images	1	22	8
Dataset	6	36	19