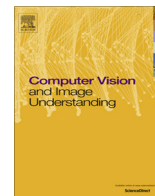




Contents lists available at ScienceDirect

Computer Vision and Image Understanding

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

Nonparametric label propagation using mutual local similarity in nearest neighbors



D. Giordano, I. Kavasidis, S. Palazzo*, C. Spampinato

Department of Electrical, Electronics and Computer Engineering, University of Catania, Viale Andrea Doria, 6, 95125 Catania, Italy

ARTICLE INFO

Article history:

Received 5 February 2014

Accepted 11 June 2014

Available online 24 June 2014

Keywords:

Big data
Nearest-neighbor
Object classification
Data-driven approaches

ABSTRACT

The shift from model-based approaches to data-driven ones is opening new frontiers in computer vision. Several tasks which required the development of sophisticated parametric models can now be solved through simple algorithms, by offloading the complexity of the task to the amount of available data. However, in order to develop data-driven approaches, it is necessary to have large annotated datasets. Unfortunately, manual labeling of large scale datasets is a complex, error prone and tedious task, especially when dealing with noisy images or with fine-grained visual tasks.

In this paper we present an automatic label propagation approach that transfers labels from a small set of manually labeled images to a large set of unlabeled items by means of nearest-neighbor search operating on HoG image descriptors. In particular, we introduce the concept of *mutual local similarity* between the labeled query image and its nearest neighbors as the condition to be verified for propagating labels.

The performance evaluation, carried out on the COREL 5K dataset and on a dataset of 20 million underwater low-quality images, showed how big data combined to simple nonparametric approaches allows to solve effectively complex visual tasks.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

In the last decade we have witnessed an explosion of the amount of visual content publicly available through the Internet. Such a huge amount of information has led the computer vision community towards a new research trend, aimed at exploiting the intrinsic variability of the data (both in terms of the number of image/video subjects and in the number of available unique portraits of each subject) to make the algorithms simpler.

As an example, let us consider the case of object classification, which, typically is tackled by training parametric models from a set of learning data, in order to make the model learn discriminative features associated to each object class [1]. This approach shows a few disadvantages, such as the need of a representative learning data set (large enough to cover the variability of possible class instances, but small enough to keep the learning time reasonable), the need to optimize the model's parameters to suit the data distribution, and the complexity of the mathematical functions employed to check whether a given input image matches a certain class (or classes).

* Corresponding author.

E-mail addresses: dgiordan@dieei.unict.it (D. Giordano), ikavasidis@dieei.unict.it (I. Kavasidis), simone.palazzo@dieei.unict.it (S. Palazzo), cspampin@dieei.unict.it (C. Spampinato).

Nonparametric approaches exploit data to reduce the complexity of managing class models. For example, with big datasets it is possible to translate the classification task to something like “assign to an image a class from its most similar images” [2,3]. The reliability of such systems lies entirely on the data: if no images exist which are similar to the query, the closest sample might be unrelated and produce a wrong result. On the other hand, this is also the reason why these approaches are becoming popular only now: so far, the absence of datasets large enough to satisfy the hypothesis of “completeness over object class and pose” had pushed researchers towards model-based solutions; now, the high availability of images and video which can be simply retrieved by feeding specific keywords to web search engines (such as Google and Flickr) is drawing more and more attention towards the application of simple and well-known algorithms (such as *k*-means clustering [4] and nearest-neighbor search [5]) to large datasets for solving complex visual tasks, such as scene recognition, object detection, image annotation and location recognition (see Fig. 1).

However, all of these approaches depends on image quality (see examples in Fig. 2, taken from [2,3]). In these applications, noise is mainly represented as *labeling* noise, whereas images are relatively clean, clear and at good resolution. However, this is not something that can be taken for granted for all applications, especially with data coming from sources operating in real-life conditions. One

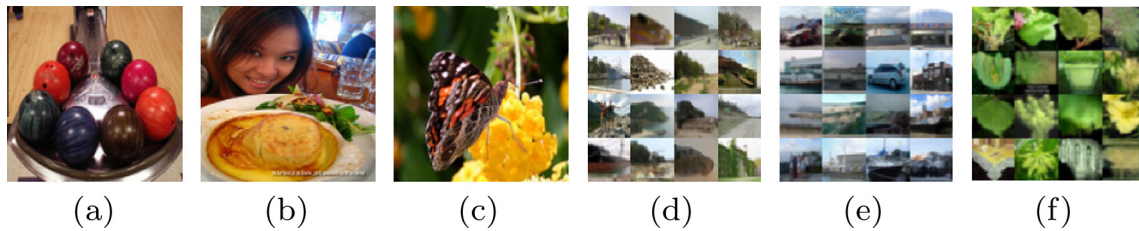


Fig. 1. Examples from the ImageNet database [3] (1a–c) and the “tiny images” dataset [2] (1d–f). Compare the quality of these images to those in Fig. 2.



Fig. 2. Examples of underwater images extracted within the Fish4Knowledge project. Top row: examples of detected fish and of the severe degradation of the overall image quality. Bottom row: examples of wrong detections.

such example is the dataset collected in the EU-funded Fish4Knowledge¹ project, which aimed at studying fish populations in their natural environment using videos taken by underwater cameras. In such scenario, weather conditions, water status, network bandwidth, storage capabilities and, most importantly, light propagation in water significantly compromise the visual level of details of detected fish, as shown in the examples of Fig. 2.

Nevertheless, the size of the resulting dataset (about 1.5 billion images) was large enough to cover a wide range of combinations of fish species, fish poses, lighting conditions, and scene backgrounds. From this consideration comes the idea behind this paper: why not try to label each image with an object class by exploiting the size and variability (albeit noisy) of the dataset itself?

The idea to achieve this objective is as follows: let us define a certain *label* to be assigned to each image; for example, this label may be, in our case, as simple as either “fish” or “non-fish”, or something more complex, such as the fish species. Since the size of the dataset allows us to assume that, given an input image, the most similar images will likely share the same label, we can exploit this assumption to label the whole dataset by expanding small sets of manually-labeled “seed” images through an iterative propagation approach, which consists in repeatedly searching for the nearest neighbors of seed images, applying a filtering approach to select the most relevant results (and reject false positives) and including the results into the corresponding label group. After each execution of this search-and-add procedure, newly-labeled images included in a given label group can be used as seeds for the next propagation runs, thus increasing the variability of the seeding sets and making the algorithm capable of including, from time to time, object instances not represented in the original labeled sets.

More in detail, this paper advances the recent state of the art on nonparametric label propagation methods in that:

- We propose an iterative label propagation method operating at large scale and extremely robust to false positives by using *mutual local similarity* between the query labeled image and the nearest neighbors as the condition to be verified for propagating labels; in particular, in contrast to image classification [2], where precision is relatively important, in our application we have a strict requirement on the precision as wrong label assignments might cascadingly affect further annotations and, therefore, the entire labeled dataset.
- We prove that simple nonparametric approaches and feature descriptors (such as quantized HoG [6]) operating on big data allows to solve extremely complex tasks such as label propagation, which is still an open problem in computer vision and machine learning.

In this work, we mean to show, at a lower scale than the size of the whole dataset generated within the F4K project, a proof of concept of the automatic labeling process described above and of the solutions adopted to deal with the difficulties that the task presents. The method was tested on a subset of 20 million images with large variability of object classes and poses to make the “completeness over object class and pose” hypothesis still valid. However, we would like to highlight that the application to fish images is only presented as a proof of concept of the approach, which is generally applicable to any similar labeling task; for the same reason, implementation choices may be changed according to different application domains (e.g. the choice of other image features than HoG).

2. Related work

Data-driven approaches generally rely on comparing high-dimensional vectors in large datasets and have been applied in several applications: image classification [7,8], object recognition [2,9,10], 3D object detection [11] and recognition [12], image retrieval [13–16], scene recognition [2,17–19], graph building and partitioning [20,21].

¹ <http://fish4knowledge.eu>.

Download English Version:

<https://daneshyari.com/en/article/527543>

Download Persian Version:

<https://daneshyari.com/article/527543>

[Daneshyari.com](https://daneshyari.com)