Pattern Recognition 56 (2016) 159–169

Contents lists available at ScienceDirect

Pattern Recognition

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

Congested scene classification via efficient unsupervised feature learning and density estimation

Yuan Yuan [a], Jia Wan [b], Qi Wang [b,*]

[a] Center for OPTical IMagery Analysis and Learning (OPTIMAL), State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, Shaanxi, PR China

[b] School of Computer Science and Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an 710072, Shaanxi, PR China

ARTICLE INFO

Article history:
Received 4 August 2015
Received in revised form
8 March 2016
Accepted 15 March 2016
Available online 24 March 2016

Keywords:
Computer vision
Unsupervised feature learning
Scene classification
Density estimation
Spherical *k*-means
Feature pooling

ABSTRACT

An unsupervised learning algorithm with density information considered is proposed for congested scene classification. Though many works have been proposed to address general scene classification during the past years, congested scene classification is not adequately studied yet. In this paper, an efficient unsupervised feature learning approach with density information encoded is proposed to solve this problem. Based on spherical *k*-means, a feature selection process is proposed to eliminate the learned noisy features. Then, local density information which better reflects the crowdedness of a scene is encoded by a novel feature pooling strategy. The proposed method is evaluated on the assembled congested scene data set and UIUC-sports data set, and intensive comparative experiments justify the effectiveness of the proposed approach.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Public traffic has become a serious problem for the development of modern cities. In order to make the transport well organized, the primary task is to know the traffic status precisely. Among various techniques enabling this function, congested scene classification is a critical one. But unfortunately, the ambiguity, variability and scale diversity in scenes make it a challenging task. In order to recognize different scenes, various methods have been proposed over the years. Most of them concentrate on the critical step of feature representation, which can be roughly divided into two categories: hand-craft features and learned features. Hand-craft features are widely used in scene classification since it is effective and interpretable [1–3]. Learned features (i.e. a transformation from raw image patches to more efficient representations) are generated by feature learning (FL) methods [4–6], which are thought to be adaptive to various situations. Because of the usefulness of FL, scene classification has been successfully used in applications such as aerial scene categorization [7], content-based image retrieval [8] and object detection [9], and has achieved notable performance in all these fields.

Congested scene classification is a specific problem of scene classification. Compared to traditional scene classification, the paper mainly focuses on the crowdedness of pedestrians and vehicles in the traffic environment, and aims to label the scene image with a level of crowdedness category. The increasing attention to this problem derives from the fact of congested traffic status. If we can get an understanding of the traffic scene by automatic analysis, the traffic management will be possible and easier [10]. Therefore, the monitoring of the transport status becomes essential. Under this circumstance, how to classify congested scenes correctly and effectively becomes a critical task.

Unfortunately, congested traffic classification is not adequately studied yet. Only a few works concentrate on crowd density measurement [11,12], traffic congestion classification [13] or crowd analysis [14]. In these tasks, the background subtraction and density estimation are the most important components. For background subtraction, optical flow is the most direct and effective method to be applied with static backgrounds. For density measurement, the local key points or individual detection are aggregated to estimate the number of objects. Nevertheless, the achieved performance is very limited.

Although many algorithms have been proposed to improve scene classification accuracy, they still have limitations when applied to congested scene classification task. On one hand, conventional approaches still have deficits. Hand-craft feature based approaches are convenient, but they are not able to generalize to

* Corresponding author.
E-mail address: crabwq@nwpu.edu.cn (Q. Wang).

http://dx.doi.org/10.1016/j.patcog.2016.03.020
0031-3203/© 2016 Elsevier Ltd. All rights reserved.

new environment. Unsupervised Feature Learning (UFL) based approaches achieve better performance, but most UFL algorithms in scene classification have many parameters to tune [15]. Moreover, the training stage is time-consuming which makes it inefficient in practical usage. On the other hand, prior information towards this particular application is not well considered. Since the paper mainly focuses on the crowdedness of pedestrians and vehicles in a scene, the representation should encode density information, which is an efficient indication of congested scenes. Unfortunately, conventional approaches do not pay attention to this important information. For example, [16] utilizes density information by simply calculating the ratio between pedestrian and road area. At the same time, most approaches separately take pedestrians and vehicles into consideration. But in the real world, pedestrians and vehicles always appear simultaneously which makes the classification results less accurate.

To address congested scene classification and alleviate problems mentioned above, a new data set which contains three different levels of congested scenes is first assembled. Subsequently, an efficient unsupervised feature learning method is exploited for low-level features extraction. Based on the obtained feature prototypes, density information is added in the image representation to help discover the congested appearances in scenes. The major contributions of this paper are summarized as follows:

1. Since there is no available data set for congested scene classification, we assemble a new data set including three types of traffic scenes: crowded scene, normal scene and open scene. This data set will provide a platform for the community and promote the research of congested scene classification with dramatically changed backgrounds. Typical exemplar images of this data set are shown in Fig. 1.
2. A more efficient approach based on spherical *k*-means and feature selection is proposed for congested scene classification. Recently, the most widely used UFL algorithm in scene classification is sparse auto-encoder (SAE) [7,17,8,5]. However, the training of SAE is time-consuming, which makes it inefficient in practical usage. Motivated by this point, this paper proposes a more efficient UFL algorithm—spherical *k*-means. It does not have any parameters and the training is faster than other UFL algorithms. Besides, instead of taking all the learned features completely like traditional treatment, we refine them to eliminate the noisy ones with a feature selection post-processing. This strategy can ensure a high-quality feature utilization.
3. Different from conventional scene classification approaches, a density estimation method is proposed to encode density information. Without detecting all pedestrians and vehicles, the proposed algorithm directly models scenes by exploiting the variations in local spatial arrangements of structural patterns captured by the learning procedure. Then a pooling operation is conducted to estimate the density according to the feature response. Therefore, our processing is more robust and reliable.

Fig. 1. Typical images in the congested scene data set. From left to right, the three columns represent crowded scene, normal scene and open scene.

The remainder of this paper is organized as follows. Related work is elaborated in Section 2. The details of the proposed approach are followed in Section 3. The performance of the proposed method is reported in Section 4. Finally, conclusion and future work are presented in Section 5.

2. Related work

In this section, the relevant works which have made great progress for decades in scene classification are reviewed. As aforementioned, literatures mainly concentrate on feature representation.

We start from low-level features [18–20], such as histogram of color/texture and power spectrum [21]. It is simple and effective for binary classification (e.g., indoor versus outdoor, man-made versus natural). However, the so called “semantic gap” between low-level features and high-level semantic labels becomes the bottleneck for further improvement. Moreover, the accuracy drops significantly when category's size becomes large.

To overcome the limitation of low-level features, bag-of-feature (BOF) based models [22–24] are proposed. Basic BOF model [25] consists of feature learning and feature encoding. For the first step, clustering algorithm, typically *k*-means, is used to generate the codebook. After that, the input image is represented by histogram of unordered codewords. This model can encode the prior information and reduce the “semantic gap”, which is significantly superior to low-level features. But the spatial clue is lost as the histogram is orderless. Besides, only a small amount of information is utilized since BOF uses hard-assignment for feature coding.

The feature encoding strategy in basic BOF model is inefficient. Thus more sophisticated feature encoding methods [26,27] are proposed to reduce the information lost. These methods can be broadly divided into two steps: feature coding and feature pooling. In the coding step, instead of hard-assignment, sparse coding (SC) is included to achieve lower reconstruction error and make the representation specialized [28]. Locality-constraint coding (LLC) which is inspired by [29] is another coding method [30]. It is smoother than SC because the similar local descriptors have similar codes by sharing bases. More extensive studies of coding methods can be found in [31]. In the pooling step, motivated by [32], Lazebnik et al. [33] proposed a spatial pooling method, namely Spatial Pyramid Matching (SPM), to encode spatial information efficiently. SPM divides the input image into spatial bins at different scales and then all histograms of visual-word in these bins are concatenated to produce the final representation. A more flexible pooling method has been proposed in [34] recently which jointly learns appearance and important spatial pooling region (ISPR). This method improves the performance by reducing the influence of false responses as the representative objects of the scene appear at several important regions with high possibility. More variants to encode spatial information in BOF can be found in [35,36].

Compared to BOF based model, object filter based models utilize higher-level semantic features [2,37,38]. Intuitively, different scenes contain different typical objects. Thus, we can use the objects and quantities of them to characterize different scenes efficiently. Motivated by this, many off-the-shelf object filters are trained to detect objects in different scenes. The responses of these

Download English Version:

https://daneshyari.com/en/article/533178

Download Persian Version:

https://daneshyari.com/article/533178

Daneshyari.com