Neurocomputing 142 (2014) 248–255

Contents lists available at ScienceDirect

Neurocomputing

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

Object categorization in sub-semantic space

Chunjie Zhang [a], Jian Cheng [b,*], Jing Liu [b], Junbiao Pang [c], Chao Liang [d], Qingming Huang [a,e], Qi Tian [f]

[a] School of Computer and Control Engineering, University of Chinese Academy of Sciences, 100049 Beijing, China
[b] National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, P.O. Box 2728, Beijing, China
[c] College of Computer Science and Technology, Beijing University of Technology, 100124, China
[d] National Engineering Research Center for Multimedia Software, School of Computer, Wuhan University, 430072 Wuhan, China
[e] Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, 100190 Beijing, China
[f] Department of Computer Sciences, University of Texas at SanAntonio TX, 78249, USA

ARTICLE INFO

Article history:
Received 11 July 2013
Received in revised form
13 March 2014
Accepted 28 March 2014
Communicated by Jinhui Tang
Available online 27 May 2014

Keywords:
Object categorization
Sub-semantic space
Structure regularized SVM
Sparse coding

ABSTRACT

Due to the semantic gap, the low-level features are unsatisfactory for object categorization. Besides, the use of semantic related image representation may not be able to cope with large inter-class variations and is not very robust to noise. To solve these problems, in this paper, we propose a novel object categorization method by using the sub-semantic space based image representation. First, exemplar classifiers are trained by separating each training image from the others and serve as the weak semantic similarity measurement. Then a graph is constructed by combining the visual similarity and weak semantic similarity of these training images. We partition this graph into visually and semantically similar sub-sets. Each sub-set of images is then used to train classifiers in order to separate this sub-set from the others. The learned sub-set classifiers are then used to construct a sub-semantic space based representation of images. This sub-semantic space is not only more semantically meaningful than exemplar based representation but also more reliable and resistant to noise than traditional semantic space based image representation. Finally, we make categorization of objects using this sub-semantic space with a structure regularized SVM classifier and conduct experiments on several public datasets to demonstrate the effectiveness of the proposed method.

1. Introduction

Object categorization refers to predicting the category of an image based on its semantic content. The state-of-the-art methods leverage the local feature based image representation technique. Local features are first extracted either by detection or dense sampling. *k*-means clustering or sparse coding is then used to generate the codebook and encode these local features. Images are then represented by bag-of-visual-words (BoW) representation and classifiers (e.g. support vector machine) are learned for object category prediction. The BoW model has been proven effective by many researchers [1–5]. However, the visual words have no explicit semantic correspondence with human perception which hinders the discriminative power of the BoW model.

To alleviate this problem, a lot of works have been done. On one hand, more discriminative and powerful features are proposed [2–5]. These well designed features capture more information and are more robust to outliers than the traditional BoW model. For example, the spatial pyramid matching (SPM) is proposed by Lazebnik et al. [2] and is widely used since its introduction. With the fast development of computational power, the explore of more powerful features for object categorization will be a hot topic in the future.

On the other hand, the use of semantic related representation is also widely studied. Semantic space based image representation is more interpretable than using local features directly for humans. These semantic spaces can be generated by latent space learning [6–8], using the training images [9,10] or generic object classifiers [11]. However, the learning of effective and robust semantic space is very hard due to the well-known semantic gap. Besides, the semantic space is often learned using all the training images of the same class. However, objects may pose large inter class variations which makes it very difficult to learn reliable classifiers for robust semantic representation. For example, the frontal view and the side view of a car are quite different. To alleviate this problem, the

* Corresponding author.
E-mail addresses: cjzhang@jdl.ac.cn (C. Zhang), jcheng@nlpr.ia.ac.cn (J. Cheng), jliu@nlpr.ia.ac.cn (J. Liu), junbiao_pang@bjut.edu.cn (J. Pang), cliang@whu.edu.cn (C. Liang), qmhuang@jdl.ac.cn (Q. Huang), qitian@cs.utsa.edu (Q. Tian).

http://dx.doi.org/10.1016/j.neucom.2014.03.059
0925-2312/

use of attribute is introduced [12–15] which helps us to improve the object categorization performance. However, the attributes have to be pre-defined. Besides, choosing the proper attributes requires experience and extensive hard work which limits its scalability for large scale applications.

Recently, the use of exemplar image for object detection [16] and categorization [17] becomes popular. The use of exemplar classifier takes the advantage of semantic space based image representation and is also more efficient and easy to train than traditional methods [6–11]. Although proven effective, not all of the exemplar classifiers are equally useful for object categorization. It would be more effective if we can choose some discriminative exemplar classifiers instead of using all of them. Besides, images of the same class often exhibit large inter-class variations which means the semantic meaning of exemplar classifiers may not be so semantic meaningful for efficient categorization.

To solve these problems, in this paper, we propose a novel object categorization method by using sub-semantic space based image representation (S^3R). First, we follow [17] and train the exemplar classifier for each training image which serves as the weak semantic similarity measurement. A visual and semantic similarity graph is constructed by combining the visual similarity and weak semantic similarity of training images. We then partition this graph to get sub-sets of images which are visually and semantically similar. Each sub-set of images is used to construct a sub-semantic space representation of images. This is achieved by learning SVM classifier which separates one sub-set of images from the others. Since we use a sub-set of images for representation, this semantic space is named as sub-semantic space. This sub-semantic space based image representation is not only more semantically meaningful than exemplar based representation but also more resistant to noise than traditional semantic space based image representation. Finally, we train structure regularized SVM classifiers with the proposed sub-semantic space for object categorization, as [17] did. We conduct experiments on several public datasets to show the effectiveness and efficiency of the proposed object categorization in the sub-semantic space method. The flowchart of the proposed method is given in Fig. 1.

The rest of the paper is organized as follows. Section 2 gives the related work. In Section 3 we show the details of sub-semantic space based image representation and give the experimental results in Section 4. Finally, we conclude in Section 5.

2. Related work

In recent years, the use of local image features has become popular and has been proven very effective for image classification. This is often done within the bag-of-visual word (BoW) framework. However, due to the well-known semantic gap, the lack of explicit correspondence between visual features and semantic concepts limits its discriminative power. A lot of works have been done to alleviate this problem which fall into two perspectives. Some tried to design more sophisticated and power features or models while others used semantic based image representation techniques.

The design of more discriminative and powerful features has been a hot topic for object categorization [2–5] for many years. To combine the spatial layout of local features, Lazebnik et al. [2] proposed a spatial pyramid matching technique. Since its introduction, the SPM based image representation is considered as a powerful tool and its effectiveness has been proven by many researchers. Yang et al. [3] found the sparse coding that is more powerful than the traditional k-means clustering based method for object categorization, especially when max pooling is used for representation images. However, the sparse coding along with the max pooling strategy is sub-optimal since negative coding parameters are ignored during image representation. To overcome this problem, Zhang et al. [4] proposed to use non-negative sparse coding instead. A part based model was proposed by Felzenszwalb et al. [5] and was widely used by researchers. Generally, a model needs more computational power if the model is more sophisticated. Although the fast development of computer industry helps us to alleviate this problem, the careful design of efficient features and models is still a very challenging problem.

On the other hand, the use of semantic related representation is also widely studied. Semantic space based image representation is more interpretable than using local features directly for humans. These semantic spaces can be generated by latent space learning [6–8], using the training images [9,10] or generic object classifiers [11]. Hofmann [6] proposed the probabilistic Latent Semantic Analysis (pLSA) technique while Blei and Jordan [7] extended it with Latent Dirichlet Allocation (LDA). However, this latent space modeling is hard to be interpreted by human being. Oliva et al. [8] organized images with three semantic axes which is determined by psychophysical experiments. To obtain more explicit semantic space, researchers also used the training concepts as well as generic objects. Rasiwasia and Vasconcelos [9] proposed to learn a low dimensional semantic "theme" from casual image annotations and used it for scene classification. Hauptmann et al. [10] studied the influence of the number of high-level concepts for reliable video retrieval and achieved good performance. Li et al. [11] proposed an ObjectBank theme which learnt generic object detectors using the images as well as the human labeled object bounding boxes from the LabelMe and the ImageNet datasets. However, the learning of effective and robust semantic space is very hard due to the well-known semantic gap. Furthermore, objects may pose large inter-class variations which makes it very difficult to learn reliable classifiers for robust semantic representation.

To alleviate this problem, the use of attribute is also introduced [12–15] which helps us to bridge the semantic gap. Farhadi et al. [12] described object categories by a set of boolean attributes such as "has hats", "near water" and built the attribute classifiers by downloading images from the internet. Lampert et al. [13] studied the problem of transfer learning by using the attribute information to detect unseen object classes. Parikh and Grauman [14] tried to build a discriminative nameable attribute vocabulary with humans in the loop which is labor intensive. To distinguish the discriminative power of each attribute for different images, Parikh and Grauman [15] tried to learn a ranking function per attribute in order to predict the relative strength of each property of images. Generally, the use of attributes helps us to boost the performance of visual applications. However, the attributes have to be pre-

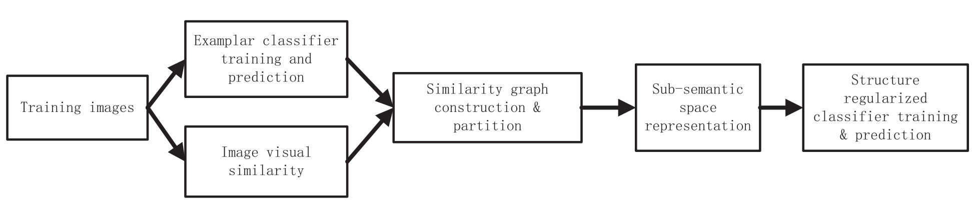

Fig. 1. Flowchart of the proposed sub-semantic space based object categorization method.