Neurocomputing 207 (2016) 105–119

Contents lists available at ScienceDirect

Neurocomputing

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

Statistical modeling for automatic image indexing and retrieval

Baopeng Zhang [a,*], Hangzai Luo [b], Jianping Fan [c]

[a] School of Computer and Information Technology, Beijing Jiaotong University, Beijing, China
[b] School of Information Science and Technology, Northwest University, Xi'an, China
[c] Department of Computer Science, University of North Carolina at Charlotte, NC 28223, USA

ARTICLE INFO

Article history:
Received 15 November 2015
Received in revised form
11 March 2016
Accepted 11 April 2016
Communicated by X. Gao
Available online 11 May 2016

Keywords:
Statistical image modeling
Statistical concept modeling
Adaptive EM algorithm
Partial similarity matching
Automatic image indexing and retrieval

ABSTRACT

In this paper, a statistical modeling algorithm is developed to achieve automatic detection of object classes and image concepts via partial similarity matching. For a given image, its statistical image model is automatically learned by using a finite mixture model to approximate the distribution of its image pixels in the 10-dimensional feature space. Such statistical image modeling process can also achieve automatic image segmentation implicitly. To achieve more precise matching between the mixture components and the local distributions of the relevant image pixels, an *adaptive EM algorithm* is developed to simultaneously select the model structure (i.e., the optimal number of mixture components) and estimate the model parameters (i.e., locations and statistical properties of the mixture components) according to the local distributions of the relevant image pixels. For a given image concept or object class of interest, its statistical concept model is automatically learned from the statistical image models for the labeled training images. Finally, similarity matching for automatic detection of object classes and image concepts is treated as a *partial model matching* problem, i.e., matching between the statistical image model for a given test image and all the statistical concept models for the object classes and image concepts of interest. Our experimental results have demonstrated that our statistical modeling algorithm can achieve very competitive results on both automatic image segmentation and classification.

1. Introduction

With the exponential growth of digital images, there is an urgent need of new techniques for automatic or semi-automatic image annotation to support keyword-based image retrieval [1,2]. One challenging issue for automatic image annotation is the semantic gap between the low-level visual features and the high-level image semantics because of huge intra-category visual diversity [17–22,36–38]. To handle the issue of huge intra-category visual diversity effectively, the image classifiers are learned from a set of labeled training images with diverse visual properties and they should be able to: (1) maintain enough intra-category visual diversity in the classifier models, so that the classifiers can easily be generalized for large amounts of unseen test images with huge visual diversity; and (2) support partial similarity matching for image classification and object detection. Another challenging issue for automatic image annotation is that images may be similar at multiple semantic levels [17–22]: the images may be semantically similar because they contain the same object class or the same image concept. Thus it is very attractive to develop the algorithms for supporting hierarchical image annotation which can simultaneously annotate the image semantics at both the object level and the concept level [17,27,30,36–38].

Statistical modeling can provide one promising and unified solution for these two challenging issues. For a given image, statistical image modeling, which focuses on finding a good matching between the mixture components and the distribution of its image pixels, can achieve automatic analysis of image contents implicitly [3–12]. For a given image concept or object class of interest, statistical concept modeling, which focuses on learning from the statistical image models for a set of labeled training images (i.e., by strengthening the mixture components which are more popular in the statistical image models for all the labeled training images and weakening the mixture components which are less popular), can provide a good solution to handle huge intra-category visual diversity effectively. Thus similarity matching for image classification and object detection is treated as an issue of *partial model matching*, e.g., partially matching between the statistical image model for a given test image and all the statistical concept models for the object classes and image concepts of interest.

In this paper, a statistical modeling algorithm is developed to enable partial similarity matching for automatic image classification and object detection. Our contribution includes the following: (1) an *adaptive EM (expectation–maximization) algorithm* is developed to

* Corresponding author.
E-mail address: bpzhang@bjtu.edu.cn (B. Zhang).

http://dx.doi.org/10.1016/j.neucom.2016.04.033
0925-2312/

simultaneously select the model structures (i.e., optimal number of mixture components) and estimate the model parameters (i.e., locations and statistical properties of the mixture components), thus more accurate modeling is achieved automatically. (2) Similarity matching for image classification and object detection is treated as an issue of partial model matching, thus our *statistical concept modeling algorithm* can handle the issue of huge intra-category visual diversity effectively. (3) Multi-level image annotation is achieved by performing *model-based object detection and scene recognition*.

The rest of this paper is organized as follows: Section 2 provides a brief review of the most relevant work; Section 3 presents a new algorithm for statistical image modeling, where an adaptive EM algorithm is introduced for model selection and parameter estimation; Section 4 introduces our statistical concept modeling algorithm by learning from the statistical image models for a set of labeled training images; Section 5 presents a new scheme for partial similarity matching; Section 6 shows the experimental results to evaluate the performance of our techniques; we conclude our paper in Section 7.

2. Related work

Some techniques have been proposed for statistical image modeling in the past [3–14]. The most prevalent framework for statistical image modeling is to use the Markov random field model [3–11]. The major limitation of a Markov random field is that its complexity increases rapidly when a larger neighborhood of image pixels is incorporated to model the local joint statistics of the image pixels. Gaussian Markov random fields have also been used for texture image modeling and segmentation [12–14]. Srivastava et al. have also proposed a universal analytical framework for statistical image modeling by considering non-Gaussian property of image statistics, i.e., high kurtosis, heavy tails, and sharp central cusps [9]. A good review for recent advances in statistical modeling of natural images is also given in [10]. With aim to pursue a unified framework for conceptualization, modeling, learning and computing visual patterns, Song-Chun has summarized the epistemological point and research threads for statistical image modeling [13]. Our review of related work consists of three parts: (a) image segmentation via statistical image modeling; (b) statistical modeling of image concepts and object classes; and (c) multi-level image annotation and classification.

To enhance the models' ability on describing diverse visual properties, the transform-domain models have been developed by performing a linear, invertible transform of the images and then modeling the transform coefficients with simpler structures [11]. To support region-based image retrieval, Carson et al. have incorporated Gaussian mixture models (GMM) for automatic image segmentation [12]. Zhuowen and Song-Chun have developed a data-driven Markov chain Monte Carlo (DDMCMC) algorithm for image segmentation which can be treated as a unifying framework for many computer vision tasks, such as edge detection, clustering, region growing, split-merge, snake/balloon, and region competition [31]. Following the same spirit of sentence parsing in speech recognition and natural language processing, Zhuowen et al. have integrated generative (top-down) methods and discriminative (bottom-up) methods for parsing images into constituent visual patterns [32]. Jiebo and Etz have also proposed an interesting approach by integrating statistical image modeling for sky detection [14], and good detection results are reported. Song-Chun has generalized some generative and discriminative techniques for statistical image modeling and applied them to conceptualize large amounts of visual patterns in natural images [13].

For visual attention modeling, Ming-Ming et al. have proposed the histogram-based GMM decomposition method for salient region detection, which focuses on a bottom-up approach for detecting the salient object regions by integrating global uniqueness and color spatial distribution [40]. In addition, Cheng et al. [41] have also developed an unsupervised segmentation algorithm, named as SaliencyCut, to support the sketch based image retrieval (SBIR) for object shape matching. To fuse multiple types of image features, multi-task low-rank sparsity pursuit (MLAP) has been developed for saliency detection [42] and image segmentation [43]. By integrating multiple features for image segmentation, MLAP [42] utilizes cross-feature information to gain better segmentation results than some traditional methods such as normalized cuts (NCut) [15] and multi-view spectral clustering [53]. More low-rank-based methods have been exploited for image segmentation and classification [44–46]. To improve the graph-based segmentation methods, the information of long range spatial connections is combined with the sparsity [54–56]. The cues of region contours, textures and boundaries have also been exploited in some existing works [39,52]. Lu and Li have developed an one-class classification approach for ASM-based (active shape model) image segmentation to classify the boundary pixels in medical images with higher accuracy [64].

Convolutional networks have recently been employed to achieve semantic image segmentation, Long et al. have exploited a new fully convolutional net (FCN) by leveraging the feature hierarchy to refine the spatial precision [47]. Chen et al. [48] have combined the fully convolutional networks with the fully connected conditional random field (CRF) to build end-to-end image segmentation map by exploiting the long-range super-pixel dependencies. Yuan et al. have developed a manifold regularized deep architecture to integrate the structure information of image features for scene recognition [63]. It is worth noting that their contributions are orthogonal to ours because our algorithm can easily leverage such deep learning features for statistical image modeling and classification.

To achieve statistical modeling of image concepts and object classes, Barnard et al. have used GMM to characterize the contextual relationships between the image concepts and the relevant image blobs (i.e., image regions with similar color or texture) [17]. By incorporating Bayesian rules for image classification, Ruofei and Zhongfei have used GMM to enable statistical concept modeling for image classification [19]. To exploit the contextual relationships or configurations among various salient image components for image semantics interpretation, Lipson et al. have proposed a simple but efficient approach for incorporating the region configurations to enhance image classification [20]. Weber et al. and Singhal et al. have also developed some quantitative approaches to incorporate the contextual relationships among various salient image components for image classification [21,22]. Li et al. have utilized the objects as the attributes of scenes to improve the task of semantic-level image recognition and classification [61]. Xuelong et al. have proposed a MAP-based (maximum a posterior) framework for scene parsing to deal with intra-category inconsistency and inter-category similarity of superpixels [62]. Jun et al. integrated click feature with varied visual features to represent image data distribution, which applied in hypergraph-based learning method [65], and ranking model to improve visual search relevance [66].

To enable multi-level image annotation, Vasconcelos has incorporated hierarchical mixture model for semantic image classification [36]. Fei-Fei and Perona have also incorporated prior knowledge to improve hierarchical image classification [37]. Based on the patch-level, object-level and scene-level information, Li-Jia et al. have proposed hierarchical generative model to integrate the tasks of classification, annotation and segmentation [60]. Li and Wang have proposed an automatic linguistic structure for image database indexing, classification and annotation [38]. Recently, Carneiro et al. have developed a supervised learning approach to enable automatic image classification and retrieval according to the image categories both at the object level and the concept level [30]. Recently, the CNNs