Neurocomputing 000 (2018) 1–14

Contents lists available at ScienceDirect

Neurocomputing

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

Model-Based segmentation of image data using spatially constrained mixture models

Can Hu, Wentao Fan*, Jixiang Du, Yuchen Zeng

Department of Computer Science and Technology, Huaqiao University, Xiamen, China

ARTICLE INFO

Article history:
Received 7 November 2016
Revised 21 July 2017
Accepted 16 December 2017
Available online xxx

Communicated by Yue Gao

Keywords:
Bayesian statistical model
Dirichlet mixture model
Spatial smoothness constraints
Variational Bayes
Image segmentation

ABSTRACT

In this paper, a novel Bayesian statistical approach is proposed to tackle the problem of natural image segmentation. The proposed approach is based on finite Dirichlet mixture models in which contextual proportions (i.e., the probabilities of class labels) are modeled with spatial smoothness constraints. The major merits of our approach are summarized as follows: Firstly, it exploits the Dirichlet mixture model which can obtain a better statistical performance than commonly used mixture models (such as the Gaussian mixture model), especially for proportional data (i.e, normalized histogram). Secondly, it explicitly models the mixing contextual proportions as probability vectors and simultaneously integrate spatial relationship between pixels into the Dirichlet mixture model, which results in a more robust framework for image segmentation. Finally, we develop a variational Bayes learning method to update the parameters in a closed-form expression. The effectiveness of the proposed approach is compared with other mixture modeling-based image segmentation approaches through extensive experiments that involve both simulated and natural color images.

1. Introduction

To analyze or extract useful contents of an observed image, it is often helpful to convert the original image data into a more explicit representation. The process of partitioning the original image into homogeneous regions is known as image segmentation, which is one of the most challenging tasks in image analysis and processing. The results of image segmentation normally affect the quality of all the subsequent processes of image analysis, such as object representation and description, and even the following higher level tasks such as object recognition or classification. However, it is quite challenging to obtain satisfied segmented results since images are often corrupted by various factors such as noise, intensity inhomogeneity, etc. Over the past few decades, numerous image segmentation approaches have been developed based on different techniques, such as graph theoretic approaches [1–4], mean shift [5,6], binary partition tree [7,8], and rate distortion theory approaches [9,10].

Except the above approaches, model-based approaches, in particular finite mixture models, have gained much attention to address image segmentation in a probabilistic manner [11,12]. A finite mixture model is composed of a finite number of basic distributions (such as Gaussian, Student's *t*, etc.) known as mixture components. It is an important statistical tool that exploits the statistical prior knowledge to model the probability density. Finite mixture models have been widely and successfully applied in large number of fields ranging from bioinformatics [13], image retrieval [14] image categorization [15], to object recognition [16]. Recently, mixture models have also been successfully applied in image emotion analysis [17,18], where Gaussian mixture models are used to represent the valence-arousal emotion labels by analyzing personalized emotion perceptions on the Image-Emotion-Social-Net data set. The parameters of the finite mixture model can be estimated using maximum likelihood (ML) or Maximum a posteriori(MAP) through the approach of Expectation-Maximization (EM). However, the EM algorithm is significantly affected by the initial values of parameters and can easily converge to a local maximum with an inappropriate initialization. In addition, when apply finite mixture models to image segmentation, their performance can be easily degraded by high levels of noise. This is mainly due to the fact that in approaches based on conventional mixture models, the spatial dependencies between nearby pixels are not taken into account, and the prior knowledge that adjacent pixels most likely belong to the same cluster is not considered.

To overcome the above shortcomings of mixture modeling-based image segmentation approaches and improve the robustness to noise, an effective approach is to impose spatial smoothness constraints to provide spatial dependencies between neighboring

* Corresponding author.
E-mail addresses: canhu@hqu.edu.cn (C. Hu), fwt@hqu.edu.cn (W. Fan), jxdu@hqu.edu.cn (J. Du), 1425131028@hqu.edu.cn (Y. Zeng).

https://doi.org/10.1016/j.neucom.2017.12.033
0925-2312/

pixels [19]. The applicability of this kind of approach can be found in many image processing tasks, such as defogging [20], denoising [21], restoration [22] and image segmentation [23]. In recent years, several image segmentation approaches based on mixture models and Markov random filed (MRF) have been proposed [24–26]. According to these approaches, by imposing the smoothness constraints on the contextual mixing proportion (i.e, the probability to assign a given pixel to a particular class), better image segmentation results can be obtained, especially when images are corrupted by high levels of noise. Nevertheless, since contextual mixing proportions are treated as a probabilistic vector and subject to the constraints that they have to be positive and sum to one, it is not straightforward to obtain a closed-form solution in the M-step of the EM algorithm. A solution to this problem is to include a gradient projection step in the M-step as in [24] with the sacrifice of computational efficiency due to additional training complexity. In addition, to make sure that contextual mixing proportions are positive and sum to one, another reparatory projection step has been introduced in the M-step to get a closed-form update solution [25]. However, this approach also requires high computational cost due to the increased complexity of the training data. Another kind of MRF, based on the spatially variant finite Gaussian mixture model was proposed in [26]. In this model, the contextual mixing proportion is explicitly modeled as a probabilistic vector which follows a Dirichlet Compound Multinomial distribution. It guarantees that the contextual mixing proportion is computed subject to the probability constraints (positive and sum to one) without introducing an extra reparatory projection step. Furthermore, spatial smoothness is imposed in this model by applying the Gauss–Markov random field. Unfortunately, the computational cost is still quite high during the process of estimating parameters due to the complex representation of the log-likelihood function.

In this paper, we propose a novel mixture model with spatial constraints for image segmentation, which has the following advantages compared with other mixture modeling-based image segmentation approaches: (1) In preference to commonly used finite mixture models such as Gaussian mixture model (GMM) [11], the proposed model is based on Dirichlet mixture model. Compared with Gaussian distribution that only allows symmetric modes, the Dirichlet distribution, which is the multivariate generalization of the Beta distribution, may have multiple symmetric and asymmetric modes. Thus, motivated by its flexibility and recent promising results in modeling non-Gaussian data [15,16], we adopt the Dirichlet distribution as the parent distribution to model image pixels directly in this work. (2) The proposed mixture model takes into account the local spatial relationship between neighboring pixels by explicitly introducing a probability prior distribution over contextual mixing proportions. As a result, the requirements that the contextual mixing proportions must be positive and sum to one are explicitly defined in the proposed Bayesian model with Dirichlet distribution without resorting to the reparatory projection step. (3) Most of the mixture modeling based approaches adopt the EM algorithm to find maximum likelihood solutions for image segmentation [24–26]. Nevertheless, the EM algorithm suffers from the problems of over-fitting and the incapacity of determining model complexity. In this work, we propose a model learning approach based on variational Bayes [27,28], such that the parameters of the proposed model and its model complexity (i.e., the number of mixture components) can be efficiently optimized in a closed form.

The contributions of this work can be summarized as follows: (1) A novel image segmentation approach is proposed based on spatially constrained Dirichlet mixture models. (2) The spatial smoothness between nearby pixels is imposed by integrating a Dirichlet Compound Multinomial distribution with the Dirichlet mixture model, in order to improve the robustness of segmentation. (3) A closed-form solution to the estimation of model parameters is derived through variational Bayes. The effectiveness of the proposed spatially constrained mixture model for image segmentation is validated through extensive simulations with both synthetic and natural color images.

The rest of this paper is organized as follows. In Section 2, the proposed Dirichlet Mixture Model with spatial constraints that takes the contextual relationship between the nearby pixels is presented. In Section 3, the variational Bayes algorithm for inferencing the parameters of our model is developed with a closed-form solution. In Section 4, experimental results of the presented model with synthetic and natural images are shown. Lastly, conclusion and directions for the future work are given in Section 5.

2. The spatially constrained Dirichlet mixture model

2.1. Finite Dirichlet mixture model

Given a D-dimensional random vector $\vec{X} = (X_1, \dots, X_D)$ subject to the constraints: $\sum_{d=1}^{D} X_d = 1$ and $0 \le X_d \le 1$, the Dirichlet distribution over $\vec{X}$ with parameters $\vec{\alpha} = (\alpha_1, \dots, \alpha_D)$ is defined by

$$\mathrm{Dir}(\vec{X}|\vec{\alpha}) = \frac{\Gamma\left(\sum_{d=1}^{D} \alpha_d\right)}{\prod_{d=1}^{D} \Gamma(\alpha_d)} \prod_{d=1}^{D} X_d^{\alpha_d - 1}, \tag{1}$$

where $\Gamma(\cdot)$ is the gamma function defined as $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} \mathrm{d}t$. The mean and covariance of the Dirichlet distribution are given by

$$\mathbb{E}[X_d] = \frac{\alpha_d}{\sum_{d=1}^{D} \alpha_l}, \tag{2}$$

and

$$\mathrm{Cov}[X_a X_b] = -\frac{\alpha_a \alpha_b}{\left(\sum_{d=1}^{D} \alpha_d\right)^2 \left(\sum_{d=1}^{D} \alpha_d + 1\right)}. \tag{3}$$

It is noteworthy that the Dirichlet distribution is a multivariate distribution which is often used as the conjugate prior of the the multinomial distribution in Bayesian statistics. However, in this work, the Dirichlet distribution is used as the parent distribution to explicitly model image pixels.

A finite mixture of Dirichlet distributions with K components can be constructed as

$$p(\vec{X}|\vec{\pi}, \vec{\alpha}) = \sum_{k=1}^{K} \pi_k \mathrm{Dir}(\vec{X}|\vec{\alpha}_k), \tag{4}$$

where $\vec{\pi} = (\pi_1, \dots, \pi_K)$ represents the set of mixing proportions with the following constraints: $0 \le \pi_k \le 1$ and $\sum_{k=1}^{K} \pi_k = 1$. $\mathrm{Dir}(\vec{X}|\vec{\alpha}_k)$ is the Dirichlet distribution of the k-th component with its own parameter $\vec{\alpha}_k$.

Assume that a set of N independent identically distributed vectors $\mathcal{X} = (\vec{X}_1, \dots, \vec{X}_N)$ are generated from the Dirichlet mixture model, the density function of $\mathcal{X}$ is given by

$$p(\mathcal{X}|\vec{\pi}, \vec{\alpha}) = \prod_{i=1}^{N} \left[\sum_{k=1}^{K} \pi_k \mathrm{Dir}(\vec{X}_i|\vec{\alpha}_k) \right]. \tag{5}$$

Recently, finite Dirichlet mixture models have been successfully adopted to solve various problems in the fields of pattern recognition and computer vision (such as image categorization, anomaly intrusion detection and object recognition, etc.) [15,16,29,30]. Unfortunately, the applicability of conventional Dirichlet mixture model is significantly limited in image segmentation. This is because conventional mixture models assume that pixels of a given image are statistically independent, and therefore the spatial correlation between nearby pixels is not taken into consideration. Another problem is that the finite Dirichlet mixture model in Eq. (5) assigns each pixel an identical mixing proportion in the