Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Image and Vision Computing 23 (2005) 587–596
image and vision COMPUTING
www.elsevier.com/locate/imavis

Bayesian inference for multiband image segmentation via model-based cluster trees

Fionn Murtagh[a,*], Adrian E. Raftery[b], Jean-Luc Starck[c]

[a]*Department of Computer Science, University of London, Royal Holloway, Egham, Surrey TW20 0EX, England*
[b]*Department of Statistics, University of Washington, P.O. Box 354322, Seattle, WA 98915-4322, USA*
[c]*SEI-SAP/DAPNIA, CEA-Saclay, Gif-sur-Yvette Cedex F-91191, France*

Received 29 January 2004; received in revised form 3 October 2004; accepted 2 February 2005

Abstract

We consider the problem of multiband image clustering and segmentation. We propose a new methodology for doing this, called model-based cluster trees. This is grounded in model-based clustering, which bases inference on finite mixture models estimated by maximum likelihood using the EM algorithm, and automatically chooses the number of clusters by Bayesian model selection, approximated using BIC, the Bayesian Information Criterion. For segmentation, model-based clustering is based on a Markov spatial dependence model. In the Markov model case, the Bayesian model selection criterion takes account of spatial neighborhood information, and is termed PLIC, the Pseudolikelihood Information Criterion. We build a cluster tree by first segmenting an image band, then using the second band to cluster each of the level 1 clusters, and continuing if required for further bands. The tree is pruned automatically as a part of the algorithm by using Bayesian model selection to choose the number of clusters at each stage. An efficient algorithm for implementing the methodology is proposed. An example is used to evaluate this new approach, and the advantages and disadvantages of alternative approaches to multiband segmentation and clustering are discussed.

Keywords: Bayesian model; Markov model; Potts; Ising; Information criterion; Quantization; Clustering; Segmentation; Multiband; Multispectral; Hyperspectral; Multichannel; Information fusion

1. Introduction

Clustering and segmentation in an image analysis context have a long history [19]. Objectives include: quantization of data values for later use with a codebook in a compression context; targeting delivery to display devices supporting small, bounded pixel data value depth; as a preliminary to object and feature detection and analysis in images; and as a basis for other image processing operations such as image registration and archiving.

We will use the terms clustering or quantization to refer to determining clusters among image grayscale or pixel values. In the case of multiband images, the grayscale pixel values are multidimensional. This simply implies that in multiband data clustering we are dealing with clustering in multidimensional space, i.e. we are dealing with a form of vector quantization. Multiband images include the case of color images, with bands associated with red, green and blue colors, or a large number of alternative color formatting schemes. As opposed to clustering or quantization, the term segmentation is used when neighborhood or spatial influence information is incorporated into the modeling. Ideally, we could impose as a *necessary* objective that all segments be spatially contiguous. In practice, we take this as a *sufficient* objective. Multiband images are also referred to as multispectral or multichannel or hyperspectral images.

In this paper, we propose a new method for multiband image clustering, called *model-based cluster trees*. This combines maximum likelihood estimation of finite mixture models with Bayesian model selection. For segmentation, a Markov neighborhood dependency model is used to include adjacency or local influence. The model-based clustering tree algorithm operates recursively on the image bands. First it clusters or segments the pixels on the basis of the first

* Corresponding author. Tel.: +44 1784 443429; fax: +44 1784 439786.
E-mail addresses: fmurtagh@acm.org (F. Murtagh), raftery@stat.washington.edu (A.E. Raftery), jstarck@cea.fr (J.-L. Starck).

0262-8856/$ - see front matter
doi:10.1016/j.imavis.2005.02.002

band. Then, using the second selected band, it clusters each of the clusters found in the first stage. Bayesian model selection is used at each stage to determine the number of clusters or segments, so that the data are used to decide adaptively the extent to which the tree is pruned.

The resulting method allows the number of quantization levels or numbers of segments to be chosen on the basis of the data. If the number of quantization levels is predetermined (see, e.g. [23]), the method can easily handle this as a special case. Given that image bands are processed in sequence, it is helpful if the image bands have some inherent order. In chromaticity/luminosity color space, such an order can make use of the fact that chromaticities convey far less perceptual information than does the luminosity (see, e.g. [32]). Such an order can be readily accommodated in our approach. In more general cases, we impose an order on image bands which will be helpful for interpretation or further processing of the clustered or segmented output.

We can readily accommodate noise in our image data. This is implied by image features taken as realizations of distributional models. Explicit noise components are incorporated into our modeling as discussed in earlier work of ours [4]. Our MR software package [17] provides multiband image noise filtering, together with compression, functionality. See also chapter 6, ‘Multichannel data’, in [30].

We can accommodate a very small number of classes (clusters or segments) for the pixels, or a large number. A small number of classes may be needed as a preliminary to a data interpretation, or high-level vision stage of the analysis. A large number of classes may be needed when high fidelity to the original image is required.

A major motivation for a cluster tree results from use of model-based clustering in cases like multiband segmentation in Earth observation [22]. Notwithstanding the Occam razor parsimony principle of a small number of clusters, it may be found that a larger number of clusters does greater justice to the data. Then, however, it may be necessary to further analyze the clusters found. A cluster tree approach is an appropriate way to do this.

The simple tree structure given by a quadtree can be valuable, in particular for permitting Markov modeling both spatially and in scale [7]. However, two problems arise with such a simple tree structure: firstly, there is a sharp discontinuity at the boundaries between quadtree cells; and secondly the quadtree is quite a crude data-driven structure.

A further motivation for our cluster tree approach is that model-based Gaussian fitting of arbitrary multiband data is often unstable and algorithmically non-robust. The reason for this is singularity brought about by the following: (i) individual clusters or segments that are of small cardinality; (ii) correlation, possibly local, between bands; and (iii) relatively ‘flat’ background that is not covered by the detector, in particular in medical imaging. Some of these issues are discussed by us in [22].

In Section 2, we describe the model-based cluster trees methodology. In Sections 3 and 4, we discuss aspects of algorithm design and properties. In Section 5, we will exemplify where the model-based tree approach is particularly important, and show how this algorithm performs exceedingly well in practice.

2. Model-based cluster trees

Our basic framework is that of *model-based clustering*, as described, for example, by Fraley and Raftery [11,12]. In this methodology, a finite mixture of normal distributions is fit to the data by maximum likelihood estimation using the EM algorithm, the number of groups is chosen using Bayesian model selection, and if hard clustering is desired, each pixel is assigned to its most likely group a posteriori. Model-based cluster *trees* produces a clustering of multivariate data by clustering on each band or dimension recursively.

We now briefly outline finite mixture modeling, Bayesian model selection, and model-based cluster trees.

2.1. Univariate finite Gaussian mixture models

In the univariate finite Gaussian mixture model, one-dimensional observations x_i are assumed to be drawn from G groups, each of which is Gaussian distributed. The gth group has mean μ_g and variance σ_g^2. Given observations $x=(x_1,\ldots,x_n)$, let γ be an unobserved $n\times G$ cluster assignment matrix, where $\gamma_{ig}=1$ if x_i comes from the gth group, and $\gamma_{ig}=0$ otherwise. Our goals are to determine the number of clusters G, to determine the cluster assignment of each pixel, and to estimate the parameters μ_g and σ_g of each cluster.

The probability density for this model is

$$f(x_i|\theta,\lambda) = \sum_{g=1}^{G} \lambda_g f_g(x_i|\theta_g), \quad (1)$$

where $\theta_g=(\mu_g,\sigma_g^2)^{\mathrm{T}}$, $f_g(\cdot|\theta_g)$ is a Gaussian density with mean μ_g and variance σ_g^2, $\theta=\theta_1,\ldots,\theta_G$, and $\lambda=\lambda_1,\ldots,\lambda_G)$ is a vector of mixture probabilities such that $\lambda_g\geq 0$, $g=1,\ldots,G$ and $\sum_{g=1}^{G}\lambda_g=1$.

We estimate the parameters by maximum likelihood using the expectation-maximization (EM) algorithm [9,16]. For its application to model-based clustering, see [6,8,15]. This is a procedure for iteratively maximizing likelihoods in situations where there are unobserved quantities and estimation would be simple if these were known. In the clustering case, the unobserved quantities are the cluster assignments given by the matrix γ.

The EM algorithm iterates between the E step and the M step. In the E step, the conditional expectation, $\hat{\gamma}$, of γ given the data and the current estimates of θ and λ is computed, so that $\hat{\gamma}_{ig}$ is the conditional probability that x_i belongs to