Segmentation of hyperspectral images from functional kernel density estimation

Laurent Delsol, Cécile Louchet

Contribute 1
Segmentation of hyperspectral images from functional kernel density estimation

Laurent Delsol, Cécile Louchet

Abstract The processing of hyperspectral images, seen as functions that link each pixel to a curve, has become crucial, in remote sensing applications for instance. Here we tackle the problem of segmentation of such images, by carefully combining image processing tools and functional statistics, namely a Potts model and a likelihood term based on functional kernel density estimation in a Bayesian framework, and consider possible extensions.

Introduction

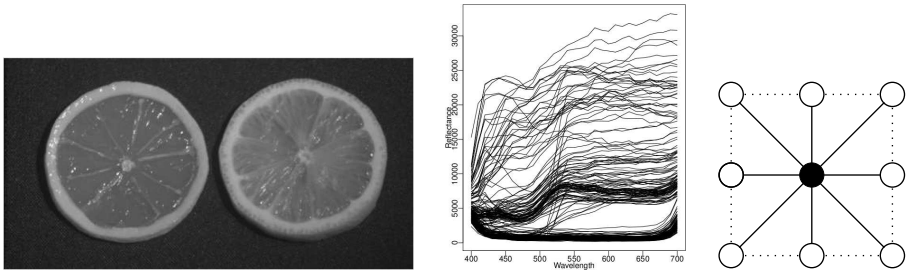
The processing of spatial data, and the segmentation of images in particular, has become more and more challenging and crucial in applications. Digital images, here viewed as mappings from a discrete set Ω (the *domain*, typically a subset of $\mathbb{Z}^2$) to $\mathbb{R}^d$, usually split into gray-level images, corresponding to $d = 1$ (a single light intensity is given for each pixel), and color images, corresponding to $d = 3$ (red, green, blue channels usually). But this is forgetting hyperspectral images (HSI) where d can be much higher (say, greater than 20): modern sensors can acquire whole spectra (224 values for each spectrum for the AVIRIS images from the NASA for instance) on reasonably sampled image grids, making the matter analysis of the sensed scene possible (see Figure 1.1 (a) and (b)).

The segmentation of an image consists in partitioning its domain Ω into disjoint regions, each region corresponding to a label, or to an object. Segmenting images (and in particular HSIs) is a crucial step before detection applications, object identification and scene understanding [1].

A classical way to segment images (with $d \leq 3$) is to optimize some simple intraclass homogeneity criterion under a region regularity prior in a Bayesian

Laurent Delsol
MAPMO, Université d'Orléans, France, e-mail: laurent.delsol@univ-orleans.fr

Cécile Louchet
MAPMO, Université d'Orléans, France, e-mail: cecile.louchet@univ-orleans.fr



(a) (b) (c)

Figure 1.1: (a) HSI from database "Fake and real lemon" (here view of a global luminance). (b) Random sample of 100 curves of the previous HSI (each with 31 reflectance measures corresponding to wavelength from 400 to 700 nm, with a 10 nm step). (c) 8-order neighbor graph structure of Ω: vertical and horizontal neighbors are counted with weight 1 and diagonal neighbors with weight $1/\sqrt{2}$.

framework [9]. Some other techniques aim to cluster temporal series with no spatial information attached to them, k-means certainly being the most popular. Now, to deal with HSIs that are curves attached to a regular grid of pixels, we must extend the previous techniques to infinite-dimensional data. Popularized by [11], functional statistics have been rapidly growing as testify [5, 8] (and references therein), and represent a natural framework for our HSI.

In order to tackle this problem of HSI segmentation, in Section **1.1** we generalize the minimal partition model (here seen as a Bayesian model) for image segmentation, to functional data, by using functional kernel density estimation. Section **1.2** is devoted to the algorithm and implementation details. In Section **1.3** we show our segmentation results on synthetic and real-world data and discuss the possible extensions.

1.1 The model

1.1.1 Notations and Bayesian segmentation

A HSI is a mapping $y : \Omega \to \mathbb{R}^d$, with d potentially high. The vector of y at pixel s is denoted by y_s. Two pixels in Ω, say s and t, will be called neighbors (denoted by $s \sim t$) when s and t are connected in the graph defined on Ω (see Figure 1.1 (c)). Hence the edges of the graph are the pairs $\{s, t\}$ such that $s \sim t$.

Segmenting y amounts to find a function $x : \Omega \to \mathcal{L}$, where $\mathcal{L}$ is a finite set of labels, such that x "fits well" with the content of y. In this paper, we will focus on the Bayesian point of view, where the desired segmentation $\hat{x}$ maximizes some probability given the original image y. Using Bayes' rule, $\hat{x}$ satisfies

$$\hat{x} = \arg\max_x P(X = x|Y = y) = \arg\max_x f_{Y|X=x}(y)P(X = x) \tag{1.1}$$

where $f_{Y|X=x}$, a conditional p.d.f. (w.r.t. some reference measure assumed to exist), and $P(X = x)$, representing the probability measure of X, are meant to be modeled in the next paragraphs.

1.1.2 Minimal partition model, independence assumption

A classical way to segment images (with low d) uses the minimal partition model [9], which we present here in a non-conventional way, but more suitable to a statistics community.

In the minimal partition model, the term $P(X = x)$ in (1.1) models the confidence in a segmentation x, forgetting about the image y; it is therefore natural to model $P(X = x)$ as a regularizing term, favoring segmentations with big connected regions and smooth boundaries. This is tracted by a generalization of Ising model to multilabel framework, known as Potts random field [10]

$$P(X = x) = \frac{1}{Z} \exp\big(- \beta \sum_{s,t \in \Omega} \gamma_{s,t} \mathbb{1}_{x_s \neq x_t} \big), \tag{1.2}$$

where Z is a universal normalizing factor, and $\gamma_{s,t}$ is the weight of the edge $\{s, t\}$ (1 if s and t are horizontal or vertical neighbors, $1/\sqrt{2}$ if they are diagonal neighbors, and 0 if they are not neighbors). The potential of this Gibbs-Markov field can be interpreted as the length of all the boundaries between the regions defined by x, weighted by a factor $\beta > 0$.

Endowing $\mathbb{R}^d$ with the usual ℓ^2-norm, the minimal partition model sets the other term $f_{Y|X=x}(y)$ in (1.1) as

$$\hat{f}_{Y|X=x}(y) = \frac{1}{Z} \exp\big(- \frac{1}{2\sigma^2} \|y - y^x\|^2 \big), \tag{1.3}$$

where for each pixel s, y^x_s is the mean of y on the region defined by x containing s. Hence (1.3) models y as a noisy version of x with white additive Gaussian noise with variance σ^2. This assumption is very strong and we relax it by the only assumption that the pixels are conditionally independent, written

$$f_{Y|X=x}(y) = \prod_{s \in \Omega} f_{Y_s|X_s=x_s}(y_s). \tag{1.4}$$

This still models y as a noise corrupting x, but now the noise at pixel s is not necessarily Gaussian nor additive and can be estimated nonparametrically.

1.1.3 How to move to hyperspectral images

We assume that the curves $(y_s)_{s \in \Omega}$ live in a functional space $\mathcal{E}$, able to enforce more regularity than $\mathbb{R}^d$. After Ferraty and Vieu's work [6] on the estimation of the regression operator on a functional variable, many kernel estimators have been adapted to the case of functional data (see [5, 7] for instance). More little has been done on curve density estimation, due to theoretical problems linked to its definition, and to the difficulty to build a (pseudo-) estimator. Here we use the seminal work of [2] stating that

$$\hat{f}(y) = \frac{1}{n R_y(h, \delta, \mu)} \sum_{i=1}^{n} K\left(\frac{\delta(Y_i, y)}{h_n} \right) \tag{1.5}$$

is an estimator of the density of Y (w.r.t. a reference measure μ), where $(Y_i)_{1 \leq i \leq n} \in \mathcal{E}^n$ is the sample, K is a kernel with support $[0, 1]$, h_n is a smoothing parameter, and δ is any semimetric defined on $\mathcal{E}$. The semimetric can be chosen according

to the nature of data ; furthermore the normalizing factor $R_y(h, \delta, \mu)$ has a closed theoretical form but remains hard to evaluate from the data. Note that recent advances on curve density estimation could also be used [3, 4].

Equation 1.5 can be used to estimate $f_{Y_s|X_s=x_s}(y_s)$ in (1.4). We assume that the density $f_{Y_s|X_s=x_s}(y_s)$ is learnt from the all (y_t) for which $x_t = x_s$, which leads to

$$\hat{f}_{Y_s|X_s=x_s}(y_s) = \frac{1}{\#\{t : x_t = x_s\} R_{y_s}(h, \delta, \mu)} \sum_{t:x_t=x_s} K\left(\frac{\delta(y_t, y_s)}{h_n}\right). \tag{1.6}$$

The proposed segmentation $\hat{x} = \arg\max_x P(X = x)\hat{f}_{Y|X=x}(y)$, with $P(X = x)$ as in (1.2) and $\hat{f}_{Y|X=x}(y)$ as in (1.4) and (1.6), makes no use of the terms $R_{y_s}(h, \delta, \mu)$ in the maximization process and is hence computable.

1.2 Algorithm

In order to compute $\hat{x}$, we could apply graph-cut methods which have the advantage of being exact. However the computational burden can be avoided by using an iterated conditional mode (ICM) algorithm initialized by a reasonable segmentation. The chosen initialization is the multivariate k-means solution, corresponding to the minimal partition model with $\beta = 0$ and a Gaussian kernel.

The principle of the ICM algorithm is the local optimization of the posterior probability: a pixel is chosen in the domain (here following a deterministic sweep) and its value is updated by the label that maximizes the posterior density, given the rest of the image. Let us recall that the number of labels is finite and it then suffices to try them all. The process is repeated until the segmentation becomes stationary; the limit is a local maximum of the posterior probability.

In practice the proposed algorithm only requires around 20-100 iterations to reach convergence.

1.3 Experiments and discussion

First we consider a synthetic HSI where a 100-point curve is drawn at each pixel obeying to a certain distribution depending only on the region's label. In the first row of Figure 1.2, the HSI is depicted as a frame, subsampled by a factor 20. In the second row, the target distribution is compared to the k-means clustering result (yielding 65.70% of misclassified pixels for 6 classes), a regularized version of the latter obtained by majority rule on weighted neighborhoods (55.19% of misclassified pixels), and then our Maximum A Posteriori approach (yielding 0.21% misclassified pixels, resp. 0.24%) for two possible initializations (k-means, resp. k-means computed on only 10 random pixels), which used an Epanechnikov kernel K, a usual ℓ^2-metric and $(h, \beta) = (0.14, 0.35)$. The third row gives the behavior of our algorithm throughout the iterations.

In Figure 1.3, our method is tested on an indoor HSI of fake and true lemon slices. Using the Epanechnikov kernel and $(h, \beta) = (6000, 10)$ gives an unsatisfying segmentation: many pixels of the true lemon's pulp are classified as fake lemon's pulp. But we can observe from the spectra that their derivatives are

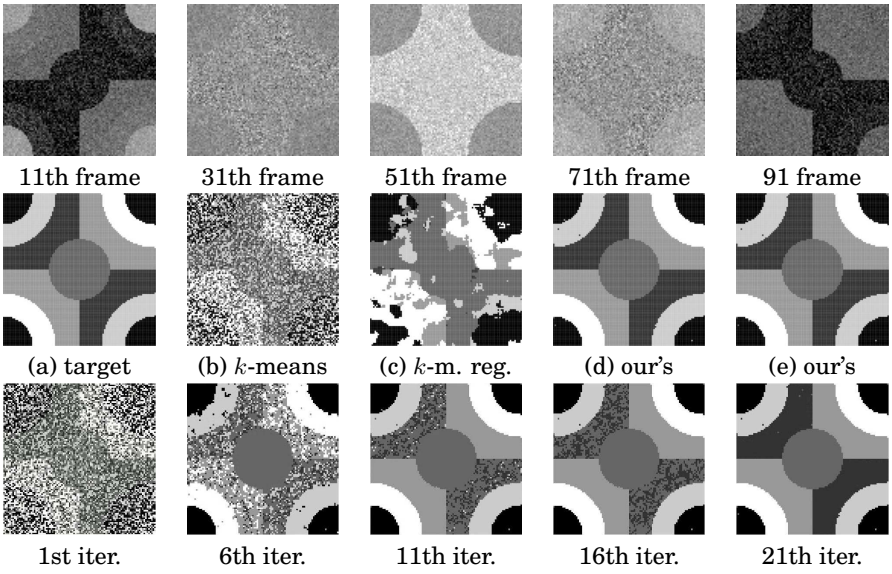


Figure 1.2: Synthetic HSI segmentation. First row: frames from the 100-frame movie representing the HSI. Second row: (a) target segmentation; (b) segmentation by k-means with 6 classes; (c) previous segmentation regularized by local majority rule; (d) our result starting from k-means (b); (e) our result starting from k-means on only 10 random pixels (similar result). Third row: initialization of our method by k-means (b) and samples of some iterations. The algorithm stops at iteration 23 where it becomes stationary.

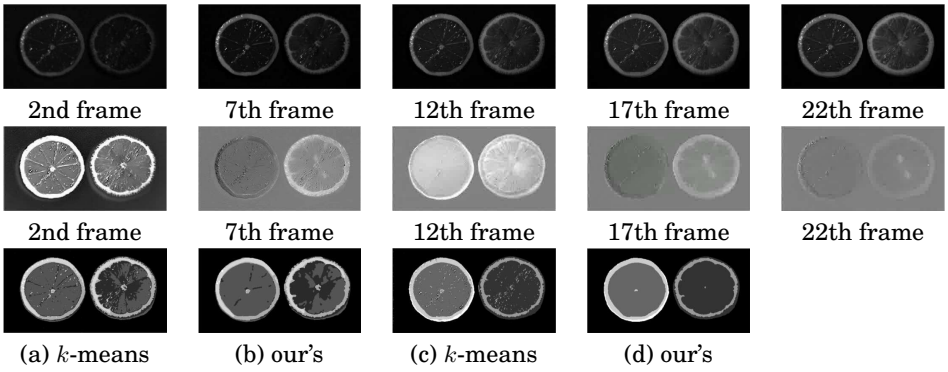


Figure 1.3: Indoor HSI segmentation. First row: frames from the 31-frame movie representing the HSI. Second row: (saturated) frames from the movie representing the HSI with the derivatives of the curves. Note that the lemon slices are visually better dissociated in this derivative representation. Third row: (a) k-means on raw data, (b) our method initialized with (a) with usual ℓ^2-metric, (c) k-means on the derivative of HSI, (d) our method initialized with (c) with $\delta = \ell^2$-metric on the first derivatives.

potentially much more discriminative than the basic spectra. In Figure 1.3 we choose for the semimetric $\delta(y_t, y_s)$ the ℓ^2-distance between the derivatives of y_t and y_s with $(h, \beta) = (20000, 10)$ and see that this semimetric is able to discriminate the true and the fake lemon almost perfectly. Thus the proposed method is flexible enough to segment natural HSIs in a non-controlled light environment.

To conclude, we proposed a relevant method to segment HSIs combining tools from functional statistics and from image processing and our numerical experiments are satisfying. The automated selection of parameters β and h is still in question, as well as some asymptotic consistency result for our segmentation. The extension of our method to gray-level image segmentation, by assigning to each pixel a curve that codes the structure of its neighborhood, is promising. Our method could also be linked to other methods in spatial functional statistics.

Bibliography

[1] G. Aubert, P. Kornprobst (2006). *Mathematical Problems in Image Processing*, Applied Mathematical Science, 147, Springer.

[2] S. Dabo-Niang (2004), Kernel density estimator in an infinite-dimensional space with a rate of convergence in the case of diffusion process. *Applied Mathematics Letters*, 17(4), pp. 381-386.

[3] A. Delaigle, P. Hall (2010). Defining probability density for a distribution of random functions. *The Annals of Statistics*, 38(2), 1171-1193.

[4] F. Ferraty, N. Kudraszow, P. Vieu (2012). Nonparametric estimation of a surrogate density function in infinite-dimensional spaces, *Journal of Nonparametric Statistics*, 24(2), 447-464.

[5] F. Ferraty, Y. Romain (2011). *Oxford Handbook of Functional Data Analysis*. Oxford University Press.

[6] F. Ferraty, P. Vieu (2002). The Functional nonparametric model and application to spectrometric data. *Computational Statistics*, 17(4), 545-564.

[7] F. Ferraty, P. Vieu (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. Springer Series in Statistics, Springer-Verlag, New York.

[8] L. Horváth, P. Kokoszka (2012). *Inference for Functional Data with Applications* (Vol. 200). Springer.

[9] D. Mumford, J. Shah (1989). Optimal Approximations by Piecewise Smooth Functions and Associated Variational Problems. *Communications on Pure and Applied Mathematics*, XLII (5): 577-685.

[10] R. B. Potts (1952). Some generalized order-disorder transformations. *Mathematical Proceedings*, 48(1): 106-109.

[11] J. Ramsay, B. W. Silverman (1997). *Functional Data Analysis*. Springer Series in Statistics.