Nonparametric clustering for image segmentation

Received 26 April 2016; | Revised 6 June 2016; | Accepted 6 June 2016

DOI: xxx/xxxx

RESEARCH ARTICLE

Nonparametric clustering for image segmentation

Giovanna Menardi*

[1]Department of Statistical Sciences, University of Padova, Padova, Italy

Correspondence
*Giovanna Menardi, Email: menardi@stat.unipd.it

Present Address
via C. Battisti 241, 35121 Padova, Italy

Summary

Image segmentation aims at identifying regions of interest within an image, by grouping pixels according to their properties. This task resembles the statistical one of clustering, yet many standard clustering methods fail to meet the basic requirements of image segmentation: segment shapes are often biased toward predetermined shapes and their number is rarely determined automatically. Nonparametric clustering is, in principle, free from these limitations and turns out to be particularly suitable for the task of image segmentation. This is also witnessed by several operational analogies, as, for instance, the resort to topological data analysis and spatial tessellation in both the frameworks.

We discuss the application of nonparametric clustering to image segmentation and provide an algorithm specific for this task. Pixel similarity is evaluated in terms of density of the color representation and the adjacency structure of the pixels is exploited to introduce a simple, yet effective method to identify image segments as disconnected high-density regions. The proposed method works both to segment an image and to detect its boundaries and can be seen as a generalization to color images of the class of watershed methods.

KEYWORDS:
Image Segmentation, Kernel smoothing, Mode, Nonparametric Density Estimation

1 | INTRODUCTION AND MOTIVATION

In the recent years, the need of analysing large amounts of image information has become relevant in several contexts and applications. Daily examples include medical diagnosis based on X-ray or magnetic resonance images, video surveillance and geographic information system applications, and image tagging. A possible goal of image analysis is the one of *segmentation*, the automatic process of identifying salient regions and single objects in an image, with the purpose of content retrieval, object detection or recognition, occlusion boundary, image compression or editing.

Digital images are created by a variety of input devices, such as cameras or scanners, and they have usually a fixed resolution, *i.e.* they are represented by a fixed number of digital values, known as *pixels*. Pixels are the smallest individual element in an image, holding quantized values that represent the brightness of a given colour at any specific location of the image. When an image is segmented, a label is assigned to each pixel, so that pixels with the same label share similar characteristics in terms of colour, intensity, or texture.

This task recalls closely the aim of cluster analysis, and thereby clustering methods have been featured as a standard tool to segment images. Within this framework, an approach which naturally lends itself to the task of image segmentation is known as *nonparametric* or *modal* clustering. According to this formulation, a probability density function is assumed to underlie the data and clusters are defined as the domains of attraction of the modes of the density function, estimated nonparametrically. This correspondence between clusters and regions around the modes of the data distribution entails several reasons of attractiveness. First, the number of clusters is an intrinsic property of the data generator mechanism, thereby

well defined, at least conceptually, and its determination is itself an integral part of the clustering procedure. Additionally, modal regions comply with the geometric intuition about the notion of clusters, also because they are not bound to any particular shape. These reasons make nonparametric clustering particularly suitable for the segmentation of digital images, as segments shall be allowed to assume arbitrary shapes and an authomatic determination of the number of segments would be desirable.

In this work the use of nonparametric clustering for image segmentation is discussed. Pixel similarity is evaluated in terms of density of the colour representation and the adjacency structure of the pixels is exploited to introduce a simple method to assess the connectedness of the modal density regions.

In the following, an overview about nonparametric clustering is provided along with its connection with methods for image segmentation. A novel modal-clustering method specifically conceived for image segmentation is proposed and discussed, and several applications illustrated.

2 | BACKGROUND

2.1 | Overview of nonparametric clustering

Modal clustering hinges on the assumption that the observed data $(x_1, \dots, x_n)'$ are a sample from a probability density function $f : \mathbb{R}^d \mapsto \mathbb{R}^+$. The modes of f are regarded as the archetypes of the clusters, which are in turn represented by the surrounding regions.

The practical identification of the modal regions is usually performed according to two alternative directions. One strand of methods looks for an explicit representation of the modes of the density and associates each cluster to the set of points along the steepest ascent path towards a mode. Any optimization method can be applied to find the local maxima of the density, such as, for instance, the *mean-shift* algorithm, early proposed by Fukunaga and Hostetler [1975], and a number of its variants [Comaniciu and Meer, 2002, Yuan et al., 2012, Carreira-Perpinan, 2008]. In this work we consider, alternatively, a second strand, which associates the clusters to disconnected density level sets of the sample space, without attempting the explicit task of mode detection. Specifically, any section of f, at a given level λ, singles out the (upper) level set

$$L(\lambda) = \{x \in \mathbb{R}^d : f(x) \geq \lambda\}, \quad 0 \leq \lambda \leq \max f$$

which may be connected or disconnected. In the latter case, it consists of a number of connected components, each of them associated with a cluster at the level λ.

While there may not exist a single λ which catches all the modal regions, any connected component of $L(\lambda)$ includes at least one mode of the density and, on the other hand, for each mode there exists some λ for which one of the connected components of the associated $L(\lambda)$ includes this mode at most. Hence, all the modal regions may be detected by identifying the connected components of $L(\lambda)$ for different λs. Varying λ along its range gives rise to a hierarchical structure of the high-density sets, known as the *cluster tree*. For each λ, it provides the number of connected components of $L(\lambda)$, and each of its leaves corresponds to a *cluster core*, i.e. the largest connected component of $L(\lambda)$ including one mode only. Figure 1 illustrates a simple example of this idea: cluster cores associated with the highest modes 2 and 3 are identified by the smallest λ larger than λ_3, while the smallest λ larger than λ_1 identifies the cluster core associated to mode 1. In some sense, the cluster tree provides a tool to inspect data with different degree of sharpness: clusters 2 and 3 are distinct, but they merge together to create a lower resolution cluster.

From the operational point of view, a few choices are required to implement the ideas underlying modal clustering. Since f is unknown, a nonparametric estimator $\hat{f}$ is employed to obtain its representation. A common choice for multidimensional data is the product kernel estimator [see, e.g., Scott and Sain, 2005]:

$$\hat{f}(x; h) = \sum_{i=1}^{n} \frac{1}{n h_1 \cdot h_d} \prod_{j=1}^{d} K\left(\frac{x^{(j)} - x_i^{(j)}}{h_j}\right) \qquad (1)$$

where $x^{(j)}$ is the $j - th$ component of x, the univariate kernel K is usually taken to be a nonnegative function centred at zero and integrating to one, and a different smoothing parameter h_j is chosen for each component. In fact, for the development of the method, it does not really matter which specific estimator is adopted, provided that $\hat{f}$ is positive and finite at the observed points.

A second choice derives from the lack, in multidimensional sample spaces, of an obvious method to identify the connected components of a level set. For these reasons the inherent literature has mainly focused on developing efficient methods for this task [*e.g.* Stuetzle and Nugent, 2010, Menardi and Azzalini, 2014].

Note that the union of the cluster cores does not produce a partition of the sample space, as regions at the tails or at the valleys of f, where the attraction of each mode is low, are initially left unallocated. However, the availability of a density measure allows for providing each unallocated observation with a degree of confidence of belonging to the cluster cores. Depending on the application, the evaluation of such confidence may be exploited to force the assignment or may result in the opportunity of fuzzy clustering schemes.

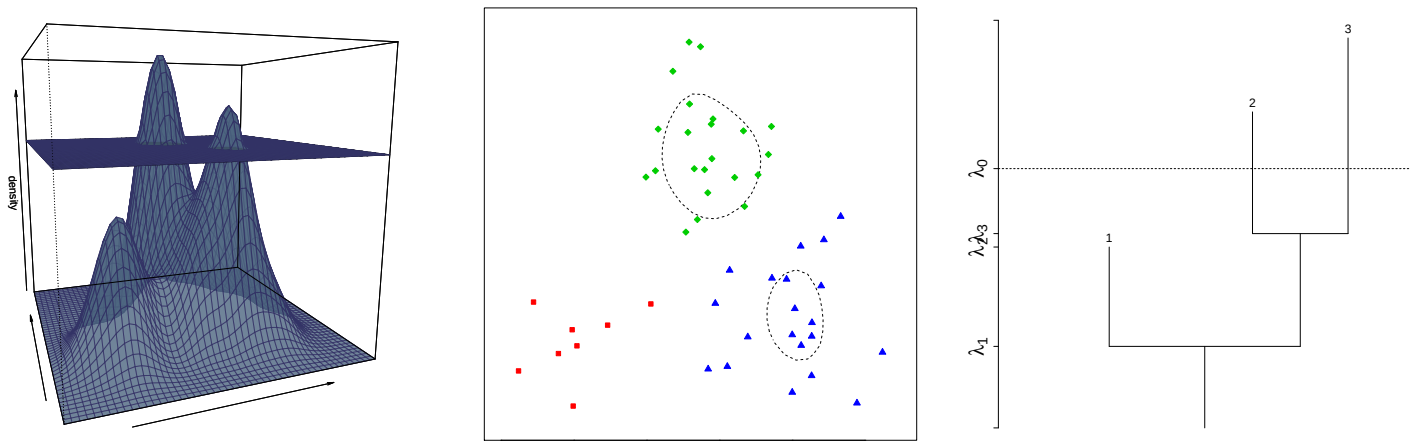

FIGURE 1 A section of a density function at a level λ_0 (left), the identified level set (middle panel), formed by two disconnected regions and the associated cluster tree, with leaves corresponding to the modes. The horizontal line is at the level λ_0 (right).

2.2 | Related works on image segmentation

Also due to the extensiveness of its applicability, several different methods have been proposed to pursue the task of image segmentation. These are broadly ascribable to two alternative routes [see, for a review, Dougherty, 2009]: *noncontextual* techniques ignore the relationships among the features in the image and assign the same label to pixels sharing some global attribute, such as the grey level or the colour brightness. *Thresholding*, for instance, compares the intensity of each pixel with a suitable threshold and associates higher values to the foreground of the image, of main interest, and lower values to the background. *Contextual* techniques, conversely, also account for pixel location or colour gradient. Within this class, *region-based methods* mainly rely on the assumption that the neighbouring pixels within one region have similar value. Boundary-based methods as *edge detection* and *active contours* build on finding pixel differences rather than similarities, to determine a closed boundary between the foreground and the background of the image. *Watershed* segmentation builds a distance map of a grey-scale image or of its gradient and considers it as topographic relief, to be flooded from its minima. When two lakes merge, a dam is built, representing the boundary between two segments.

Within the framework of clustering methods, $K-$means clustering is diffusely adopted for image segmentation, perhaps due to its simplicity. However, a few severe limitations prevent its effectiveness. First, $K-$means clustering is known to produce sub-optimal solutions as it highly depends on the initialization of the centroids. Additionally, it requires a prior specification of the number of clusters. In image segmentation this operation is undoubtedly easier than in other clustering applications. On the other hand, the need of human intervention vanishes the effort to automate the segmentation procedure. Finally, K-means is known to be biased toward the identification of spherical clusters, which can be restrictive in image data where segments may assume arbitrarily odd shapes.

While nonparametric clustering is rarely mentioned as the underlying approach to perform image analysis, it features some connection with a number of segmentation algorithms. By exploiting some notions from differential geometry, Chacón [2015] relies on Morse theory to provide an elegant formalization of the notion of modal cluster which closely recalls the ideas underlying watershed segmentation: intuitively, if the density underlying the data is figured as a mountainous landscape, and modes are its peaks, clusters are the âĂŸregions that would be flooded by a fountain emanating from a peak of the mountain rangeâĂŹ. Furthermore, the simple *thresholding* approach can be interpreted as a naive, single-λ, implementation of the density level set formulation above mentioned, where grey intensities are employed as a measure of density. Gradient ascent algorithms in the guise of mean-shift clustering are also sometimes applied for image segmentation [Tao et al., 2007]. Similar instruments are also at the basis of active contours models, where a suitable measure of energy is iteratively minimized by a gradient descent algorithm to identify the segment contours.

As a further link, even when applied to different goals, image analysis and nonparametric clustering share several tools: an example is provided by spatial tessellation as the Voronoi or Delaunay diagrams, which have been used in nonparametric clustering to identify density level sets connected components [Azzalini and Torelli, 2007] and are frequently employed in image analysis for thinning and skeletonization.

3 | A NONPARAMETRIC METHOD FOR IMAGE SEGMENTATION

3.1 | Description of the procedure

Let $\mathcal{I} = \{p_1, \dots, p_n\}$ be a digital image, where the ordered set of pixels $p_i = ((x_i, y_i), z_i), i = 1, \dots, n$, is described by the pair (x_i, y_i) denoting the coordinates of the pixels location, and by the vector z_i denoting the adopted colour model, *e.g.* $z_i = (z_i^{(r)}, z_i^{(g)}, z_i^{(b)})$ in the RGB colour representation [Soille, 2013]. In greyscale images, z_i is a scalar quantifying the grey intensity.

The particularization of nonparametric clustering in the framework of image analysis requires a density function to be defined at the pixels. A sensible choice builds $\hat{f}$ based on the color coordinates z_i. The specification of the (1) for an RGB color model is:

$$\hat{f}(z) = \frac{1}{nh_r h_g h_b} \sum_{i=1}^{n} K\left(\frac{z^{(r)} - z_i^{(r)}}{h_r}\right) K\left(\frac{z^{(g)} - z_i^{(g)}}{h_g}\right) K\left(\frac{z^{(b)} - z_i^{(b)}}{h_b}\right). \quad (2)$$

For example, if the Uniform kernel $K(u) = \frac{1}{2}\mathbf{1}_{\{|u|<1\}}$ is selected and h_j tends to zero, each pixel is provided with a density proportional to the frequency of its colour within the image. While of less immediate interpretation, a similar result holds with different kernel functions or different nonparametric estimators. As it will be discussed in Section 3.2, an alternative would also account for the spatial coordinates of the pixels.

Once that the colour density has been estimated, the associated upper level sets

$$\hat{L}(\lambda) = \{p_i \in \mathcal{I} : \hat{f}(z_i) \geq \lambda\} \qquad 0 \leq \lambda < \max \hat{f}$$

are easily determined for a grid of λ values. Next step is the identification of the connected components of the $\hat{L}(\lambda)'$s. Unlike the above mentioned case of clustering data on $\mathbb{R}^d$, where the identification of connected regions is ambiguous, the notion of connectedness is (almost) univocally determined in the image framework, due to the spatial structure of the pixels. This justifies the procedure here proposed, which builds on the level set formulation of nonparametric clustering but naturally exploits the adjacency structure of the pixels to identify the connected components of the modal regions. For a given λ, the connected components of $\hat{L}(\lambda)$ are approximated as follows:

(i) for each pixel, identify the adjacent pixels forming its *4-neighbourhood*, *i.e.* a central pixel has four connected neighbours - top, bottom, right and left.

(ii) approximate the connected components of $\hat{L}(\lambda)$ by the union of adjacent pixels in $\hat{L}(\lambda)$.

For varying λ, the procedure described so far creates K groups of pixels ℓ_k $(k = 1, \dots, K)$, which we call, in analogy with the clustering problem, (segment) *cores*, and it leaves a number of pixels unlabeled. Depending on the application at hand, we can either decide to force their assignment to the existing segment cores or to leave these pixels unallocated. In fact, a peculiar aspect is that the unlabelled points are not positioned randomly in the image, but are inevitably on the outskirts of the existing segment cores. As will be illustrated in the Section 4, unallocated pixels include (or correspond to) the contours of the segments.

The possible allocation of the unlabelled pixels to the existing groups is essentially a classification problem that may be faced according to a wide range of choices. To remain within the same kind of approach pursued so far, and consistently with the purpose of identifying segments as connected sets

Algorithm 1 Nonparametric density-based segmentation: main steps of the procedure

Require: $h = (h_r, h_g, h_b)$; ϵ All $\in$ {TRUE, FALSE} (set All:= TRUE to assign a label class to all the pixels;)

Identify the pixels adjacent to $p_i, i = 1, \dots, n$
Compute $\hat{f}(z) = \sum_{i=1}^{n} \frac{1}{nh_r h_g h_b} \prod_{j=\{r,g,b\}} K\left(\frac{z_l^{(j)} - z_i^{(j)}}{h_j}\right)$, $z \in \{z_i\}_{i=1,\dots,n}$
while $0 \leq \lambda \leq \max \hat{f}$ **do**
    identify $\hat{L}(\lambda) = \{p_i : \hat{f}(z_i) \geq \lambda\}$;
    find the connected components of $\hat{L}(\lambda)$ (as the union of adjacent pixels in $\hat{L}(\lambda)$)
    $\lambda := \lambda + \epsilon$
end while
Build the hierarchy of the connected components of $\hat{L}(\lambda)'s$ and obtain the cluster tree
Denote core pixels as p_c and unallocated pixels as p_u
Assign the label class $\ell_c \in \{1, \dots, K\}$ to each core pixels p_c
if All **then**
    while $\{p_u\} \neq \emptyset$ **do**
        for all p_u **do**
            compute $\hat{f}_k(z) = \sum_{c:\ell_c=k} \frac{1}{nh_r h_g h_b} \prod_{j=\{r,g,b\}} K\left(\frac{z^{(j)} - z_c^{(j)}}{h_j}\right)$, $k = 1 \dots, K$
            set $k_0 = \text{argmax}_k r_k(p_u) = \frac{\hat{f}_k(z_u)}{\max_{l \neq k_0} \hat{f}_l(z_u)}$
            if $\exists\, p_c$ such that p_c is adjacent to p_u and $\ell_c = k_0$ **then**
                assign the label class $\ell_c := k_0$ to p_u
            end if
        end for
        update $\{p_u\}$ and $\{p_c\}$
    end while
else
    set $\ell_u := 0$
end if
RETURN: $\ell_1, \dots, \ell_n$

of pixels, we propose the following idea to classify an unallocated pixel p_u: compute the K estimated densities $\hat{f}_k(z_u)$, each based on the pixels already assigned to the k^{th} core only $(k = 1, \dots, K)$; then, set

$$k_0 = \text{argmax}_k r_k(p_u) = \frac{\hat{f}_k(z_u)}{\max_{c \neq k_0} \hat{f}_c(z_u)} \quad (3)$$

and assign p_u to the group with label k_0 provided that at least one of the pixels already assigned to the k_0^{th} segment core is adjacent to p_u. The operational implementation of this idea is here performed in a sequential manner, as detailed in the pseudo-code above along with the main steps of the whole segmentation procedure.

3.2 | Discussion

Since the procedure illustrated above accounts for both the colours and the connectivity of the image patterns, it emulates, in some sense, the behaviour of the human eye, which instinctively, perceives different segments in a picture as either disconnected set of pixels or image patterns with a diverse colour intensity. A simple illustration of this latter aspect is withnessed by the greyscale image in Figure 2 . Even if the grey intensities of the foreground (the square) and the background are similar, the density estimator (2) perfectly distinguishes the two density levels and the isoline identifies the contours of the foreground segment. Conversely, with respect to the former aspect, a major limitation of nonparametric clustering, in principle inherited by the proposed segmentation procedure, derives from the definiton of mode itself, which requires a separating âĂŸgapâĂŹ between dense patterns. In Figure 2 , the density shows itself like a squared hole, and there is no lower density area between the background and the foreground. This prevents $\hat{L}(\lambda)$ to be a disconnected set for any λ, which would guarantee the identification of two segments. This behaviour is somewhat paradoxal, as the neater the image, the less ideal the setting for the procedure to work effectively: within an image, dripped contours of a segment, indeed, manifest themselves as small changes of colours at the borders with respect to the interior. Since the perimeter of a shape is always smaller than its area, and the density of a pixel is positively associated with the frequency of its colour, dripped contours would guarantee that the color density along the contour of a segment is lower than its inner density, and hence a valley would arise between a segment and its background. In fact, the considered example has been built ad hoc by setting the grey intensity for each pixel. In practice, many images have segment contours not defined with such neatness, no matter what the image resolution is. This is especially true with segments having either curve or sloped contours, since the shades of colours along the border of the segments allow to prevent a sawtoothed rendering. See Figure 3 for an illustration.

When the image does not features itself with dripped contours, it is possible to overcome the issue of lacking valleys in the density by introducing some blurring of the image. To this aim, given that the identification of pixel neighbours is required anyway for the identification of disconnected regions, a simple strategy is to replace each pixel value with an average of the values in its neighbourhood (in fact, a quite common practice prior to thresholding segmentation is to smooth an image using averaging or median mask. See Dougherty, 2009, §10.2.1).

A further, somewhat related, issue concerns the choice of the density measure. While we choose to build $\hat{f}$ based on the color intensities only, an alternative route would consists in exploiting the whole available information in terms of both the colour coordinates and the spatial coordinates, *i.e.*:

$$\hat{f}(z) = \frac{1}{n \prod_j h_j} \sum_{i=1}^{n} K\left(\frac{z^{(r)} - z_i^{(r)}}{h_r}\right) K\left(\frac{z^{(g)} - z_i^{(g)}}{h_g}\right) K\left(\frac{z^{(b)} - z_i^{(b)}}{h_b}\right) K\left(\frac{x - x_i}{h_x}\right) K\left(\frac{y - y_i}{h_y}\right). \quad (4)$$

This way of proceeding would also overcome the above mentioned problem of lacking valleys at the borders of the segments: in (2), the largest contribution to the density of a generic pixel is provided by all the pixels having similar colours; conversely, if also the spatial coordinates are involved in $\hat{f}$, the density of a generic pixel depends on pixels whith similar colours *and* close spatially. Hence, at the borders of a segment, where part of the adjacent pixels have a different colour, the density turns out to be lower than the interior pixels (see Figure 4 for an illustration). While this behaviour is desirable for the purpose of segmentation, the interpretation of $\hat{f}$ in terms of colour frequency fails and, indeed, a higher computational effort is required. Additionally, some empirical work not included in the manuscript has proven that estimating f via the (4) results in oversegmenting the image, hence using the only colour coordinates to estimate density is overall preferable.

In fact, a further aspect concerning the estimation of $\hat{f}$ needs to be accounted for, concerning the selection of the smoothing vector h [see, *e.g.* Silverman, 1986]. In clustering this is not as critical as it is in density estimation, since a rough indication about the high density location may suffice. However, this is certainly an issue to be tackled, and will represent object of empirical investigation in the next section.

4 | EMPIRICAL WORK

In this section, some examples are presented to illustrate the proposed procedure and to evaluate its effectiveness in identifying the salient regions for varying characteristics of the image: greyscale and multichannel images are considered; the latter ones have been selected for featuring either shaded and neated colours, and possible highlighted contours; both convex and nonconvex shaped segments have been analyzed. The procedure is compared with the performance of some competitors:

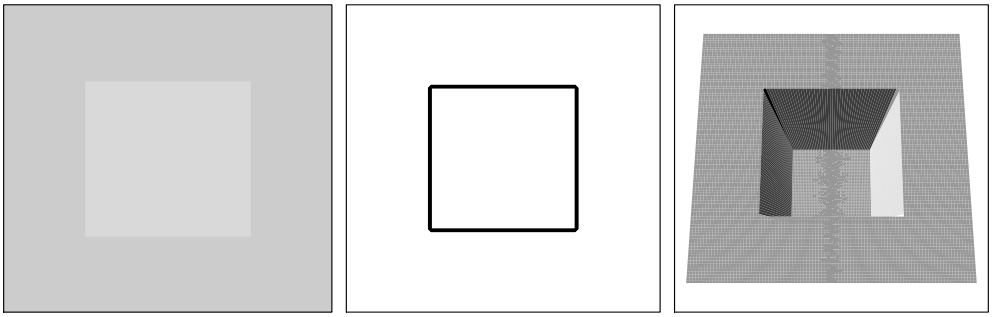

FIGURE 2 A simple grey-scale image (left), and the estimate of color density superimposed to the spatial coordinates to the pixels, represented both as the level set density (middle panel) and as a perspective plot (right panel).

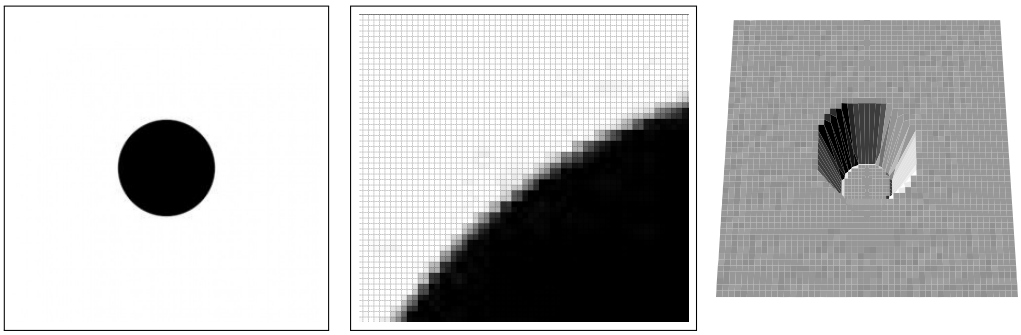

FIGURE 3 A simple grey-scale image (left), and a zoommed detail (middle), showing that when a segment has either curve or sloped contours, the colour is ripped along the border, to prevent a sawtoothed rendering. Due to this feature, the perspective plot of the density estimate, based on the (2) highlights a valley at the border of the foreground (right).

as a benchmark methods, K-means clustering is considered, as well as thresholding based on the Otsu algorithm [Dougherty, 2009, §10.2.1]. The former method has been given a head start by setting the number of clusters to its true number, as it is intuitively perceived by the author. Admittedly, the choice is not always obvious, especially for photos or, in general, shaded-color images. The latter algorithm assumes that the image contains two classes of pixels -black/white- grossly corresponding to two modal regions of the histogram built on the grey intensities. It calculates the optimum threshold separating the two classes based on the minimization of the intra-class variance. Althought it is designed for greyscale images only,

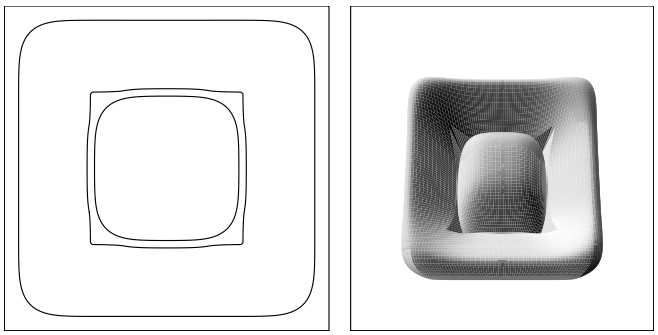

FIGURE 4 Density estimate superimposed to the greyscale image in Figure 2 when $\hat{f}$ is based both on the colours and the spatial coordinates: level set representation (left panel) and perspective plot (right panel).

thus requing a prior preprocessing of multi-channel immages, it has been considered for comparison as it represents a rought, yet effective, binary variant of the proposed procedure.

A further aim of the empirical work is to investigate the use of the density function (2) as a tool to identify the main features of an image: questions of interest are its ability to detect edges of the segment and its sensitivity to different amount of smoothing; for the considered analyses the bandwidth has been selected to the asymptotically optimal solution for data following a Normal distribution (in the following, $h = h_N$). Of course this assumpion does not hold since multi-segment images have, in principle, a multimodal density. However, this rule of thumb has resulted quite effective in clustering applications and has been then used in this analysis as a default. For the images considered as the most challenging for segmentation, the sensitivity to different choice of the bandwidth has been also evaluated.

Additionally, the ability of the cluster tree to identify a hierarchy of cluster merging that is meaningful with respect to the image has been tested. Finally, the allocation of low-density pixels, not belonging to the segment cores has been observed and commented.

All the empirical work has been performed in the `R` environment [R Core Team, 2018]. Images have been handled via the package `EBImage` [Pau et al., 2010] while the segmentation routines have been built as suitable adjustments of the clustering routines available within the package `pdfCluster` [Azzalini and Menardi, 2014].

The left panels of Figures 5 , 6 , 7 , 8 display four examples of greyscale images, selected for the analysis due to their different characteristics and different degrees of difficulty in segmentation. Multichannel images are on the other hand displayed in the left panels of Figures 9 , 10 , 11 , 12 .

Both the benchmark procedures work satisfactorily, in greyscale and multichannel images. Thresholding is intrinsecally limited, as it identifies two segments only by construction. It is, nevertheless, able to reconstruct the broad features of the image, even when it is particularly challenging. See, e.g. the famous Einstein's grimace, in Figure 8 , not easy at all to be segmented, yet very well recognizable even in the binary reconstruction via thresholding. A self-evident limitation occurs when similar color shades characterize adjacent segments, since the dichotomous choice between black or white segments unavoidably determine a corrupt reconstruction of the image. See Figure 5 for an example of this behaviour.

Despite its semplicity, $K-$means behave very well, being able to reconstruct accurately most of the images where the segment distiction is unarguable (e.g. Figures 5 , 6 , 9 , 10). On the other hand, the procedure requires the number of segments to be known in advance, and a remarkable head start has been then granted by providing the true number of segments as an input. In some applications such number might be known a priori; consider, for example, the hand X-ray image (Figure 7), where the number of segments can be set based on anatomy knowledge. If the purpose of medical image segmentation would be to attempt a first automatization of the diagnostic process, setting the number of segments to its 'normal' value, would prevent detection of fractures or anomalies, thus going in the opposite direction. A further feature of K-means is its complete noncontextuality; since distances from the cluster centroids are computed on the basis of the color only, disconnected segments might be assigned to the same cluster just for sharing a similar colour, disregarding the adjacency structure. Depending on applications, this characteristic may be sensible or not. For instance, in Figure 9 and 10 , assigning head and body of the subjects to the same segment is particularly consistent. On the other hand, a complete noncontextuality may lead to the identification of meaningless clusters, as it happens for Figure 8 whose segmentation by $K-$means results in a pixelated image.

The performance of the proposed nonparametric procedure are generally satisfactory when applied to both multichannel and greyscale images. The procedure does not result challenged by the need of distinguishing contours of assorted segments, both for size and nonconvex shapes, as especially evidenced by Figure 8 , 9 , 11 . It is especially able to identify segments as connected regions characterized by uniformity of color, but performs well also when applied to image featured by shaded colors. On the con's side, the procedure is somewhat sensitive to perceive color differences even when they are not distinguishable with the unaided eye at once, thus resulting in oversegmenting the image. As the method mainly hinges on the density, which is built on the image colors, it is in principle framed within the class of the noncontextual algorithms. However, it takes in some information about the spatial relationship between the pixels since each segment is, by construction, a (high density) connected set which is disconnected from the other segments. This characteristic prevents pixeled segmentations like $K-$means and, on the other hand does not allow the identification of unique segments as disconnected regions sharing the same color (see Figure 9 , where Bart's body and head are classified as different segments).

Note that most of the segmented images include a number of black-coloured regions, corresponding to unallocated pixels, and tiypically located at the borders of the segments. These are associated to low-density areas evaluated in the second step of the segmentation procedure. In those cases, none of the pixels already assigned to the segment presenting maximum density (3) is adjacent to the ones under evaluation, i.e., their colour is not similar to the colour of any other adjacent segment. In the analysis, these pixels are left unallocated, to

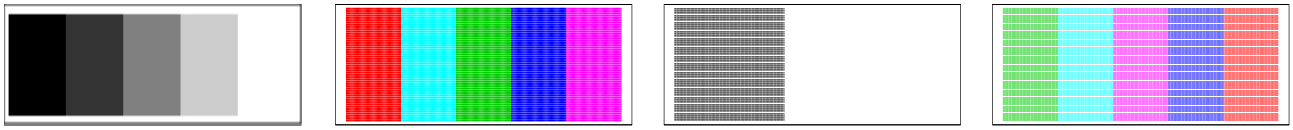

FIGURE 5 Original image (I), result of the image segmentation based on K-means clustering (II), Otsu-thresholding (III), proposed procedure (IV). Segments have been assigned arbitrary colors, except for the thresholding segmentation, where segments are either black or white by construction.

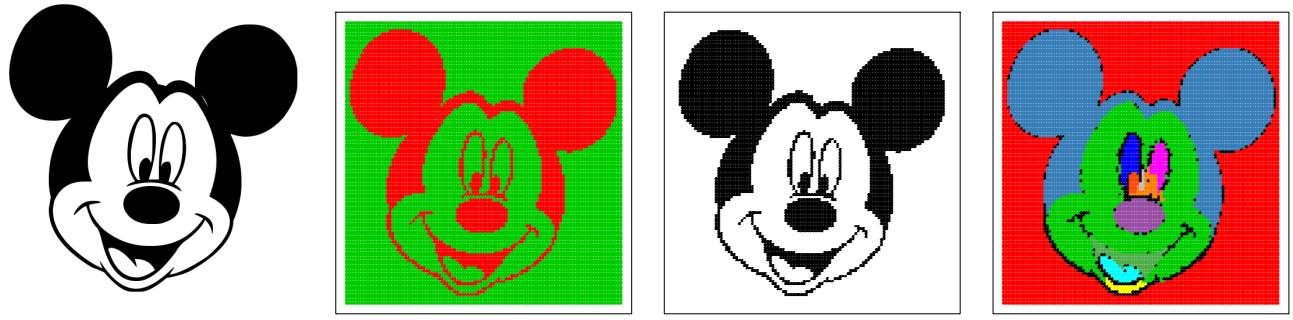

FIGURE 6 Cfr. Figure 5

enlighten the low degree of confidence in their classification. However, the issue can be worked around by either forcing the assignment to the highest density segment, disregarding spatial adjacency, or creating a further segment. Related to this aspect, and depending on subject-matter considerations, the procedure allows the opportunity of not allocating pixels that are not belonging to the cluster cores. The values of r_k provide a degree of confidence in the allocation, in the guise of fuzzy clustering schemes. This is especially useful in all the images where colours are homogeneous and segments well separated, as unallocated pixels mostly identify the boundaries of the segments (first panel of Figures 13 to 20).

With regard to the cluster tree (second panel of Figures 13 to 20), it works effectively with multichannel images in establishing a hierarchy of cluster-merging which can be associated to different levels of resolution. In the Bart image, for example, clusters that are kept separated due to small color differences (see the face and the neck) are the first to be aggregated through the cluster tree. The same does not hold for greyscale images, where the cluster tree usually cannot establish a meaningful

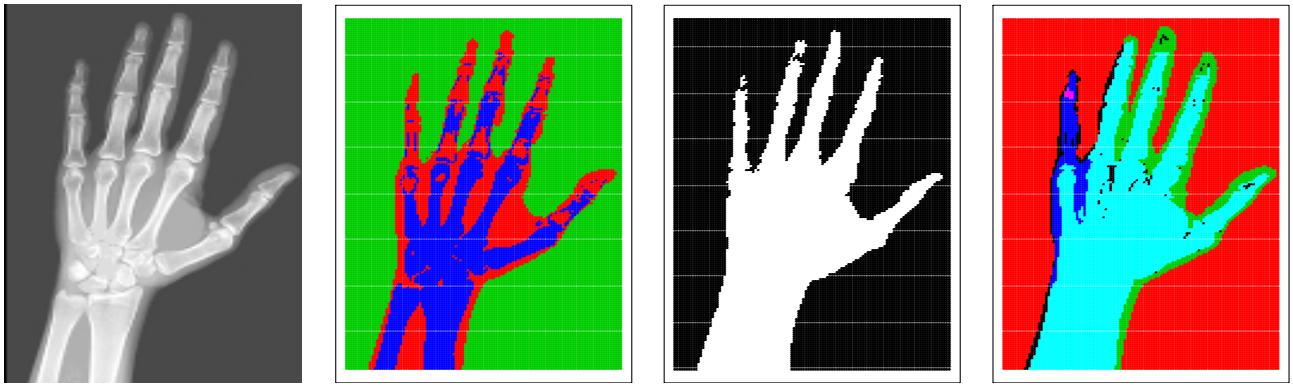

FIGURE 7 Cfr. Figure 5

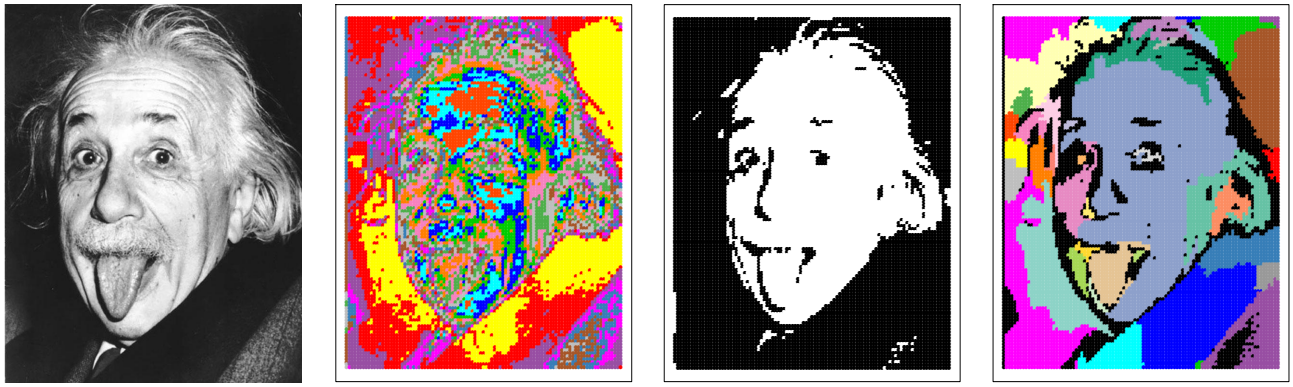

FIGURE 8 Cfr. Figure 5

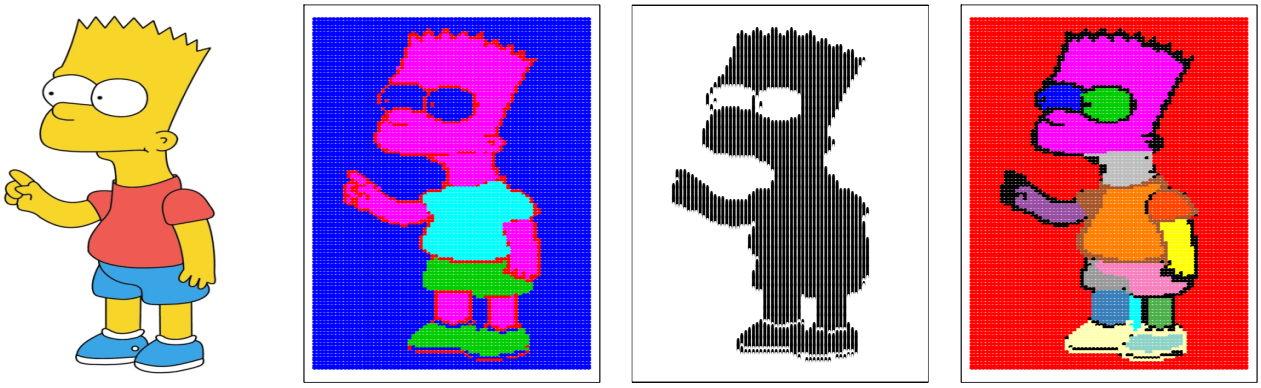

FIGURE 9 Cfr. Figure 5

hierarchy of the segments. On the other hand, in those situations also the human eye is challenged to aggregate clusters without resorting to subject-matter considerations and on the basis of the color only.

In general, the density function results an effective tools to identify the main features of the images, and density contours work well as edge detectors of the segments. See the third panel of Figures 13 to 20 . In order to understand the sensitivity of the procedure to different amounts of smoothing, the selected

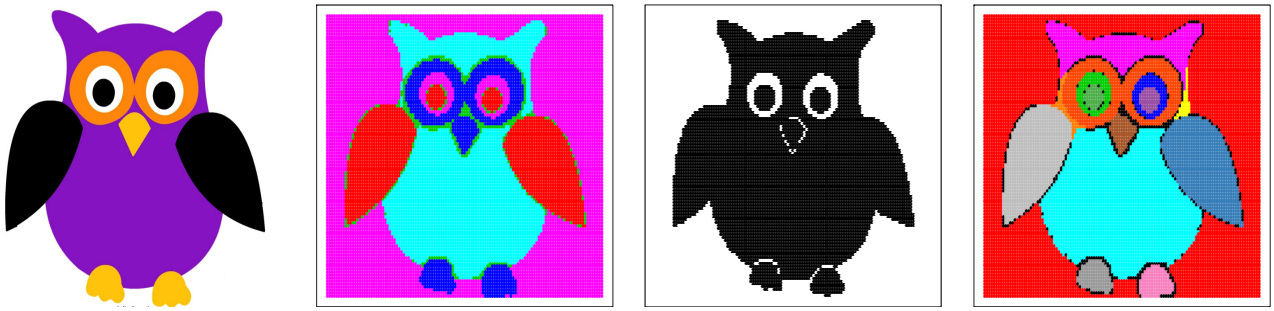

FIGURE 10 Cfr. Figure 5

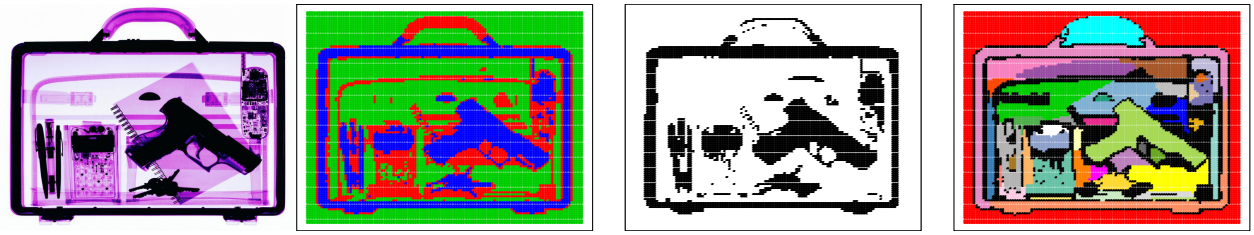

FIGURE 11 Cfr. Figure 5

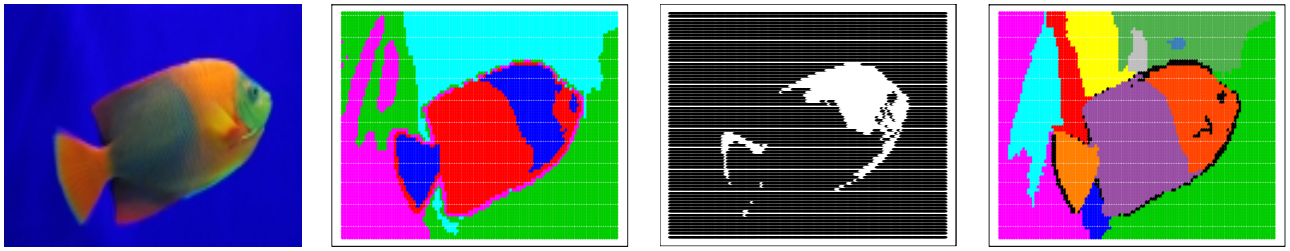

FIGURE 12 Cfr. Figure 5

hs have been varied in the range $[0.7h_N, 1.3h_N]$. Results deriving from the use of different hs are illustrated in the fourth panel of Figures 15 , 16 , 19 , 20 , and correspond to either the minimum or the maximum h for which the segmentation changes remarkably with respect to the use of h_N. In general, the segmentation is quite robust to variation of h. This is especially true for simple images with neet contours. Greyscale raws, Bart Simpson, and the owl images produce comparable segmentations over the whole range of considered values (for this reason results are omitted). As expected, more challenging images tend to be oversegmented for small h as seen, for instance, in the fourth panel of Figure 15 , where undersmoothing entails the identifications of distinct segments for the phalanxes. A large value of h, on the other hand, smoothes the density and results in segment aggregation, as seen in the last panel of Figures 16 , 20 . In fact, this monotonicity in the number of detected segments for varying h does not always occur in the application at hand, due to the use of an adaptive, finite grid of λ values to scan the density.

5 | CONCLUDING REMARKS

Image segmentation is a complicated task whose implementation cannot, in general, leave aside subject-matter considerations. All this considered, the proposed procedure is framed halfway between contextual and noncontextual segmentation algorithms, and may be then applied to a variety of situations. It can be either applied fully automatically, or be richly customized, depending on the goals of the segmentation. It is provided with some useful tools that may integrate the output of segmentation, as an estimate of the density of the pixels, which may be used to determine the degree of confidence about the segment allocation, and the cluster tree, which allows for displaying different levels of resolution of the segmentation itself.

References

A. Azzalini and G. Menardi. Clustering via nonparametric density estimation: The R package pdfCluster. *Journal of Statistical Software*, 57(11):1–26, 2014. URL http://www.jstatsoft.org/v57/i11/.

A. Azzalini and N. Torelli. Clustering via nonparametric density estimation. *Statistics and Computing*, 17:71–80, 2007.

M.A. Carreira-Perpinan. Generalised blurring mean-shift algorithms for nonparametric clustering. In *CVPR*. IEEE Computer Society, 2008.

J. E. Chacón. A population background for nonparametric density-based clustering. *Statistical Science*, 30(4): 518–532, 2015.

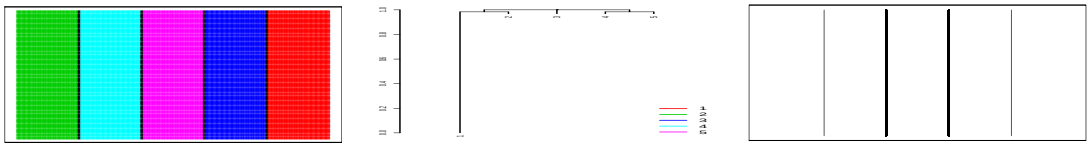

FIGURE 13 Estimated density level sets superimposed to the image (left), cluster tree (right panel) and cluster cores of the segmentation in the top left panel of Figure 5 .

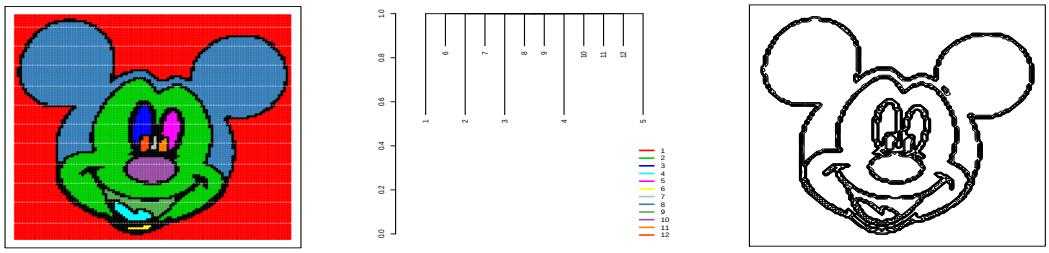

FIGURE 14 Estimated density level sets superimposed to the image (left), cluster tree (right panel) and cluster cores of the segmentation in the top left panel of Figure 6 .

D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002.

G. Dougherty. *Digital image processing for medical applications*. Cambridge University Press, 2009.

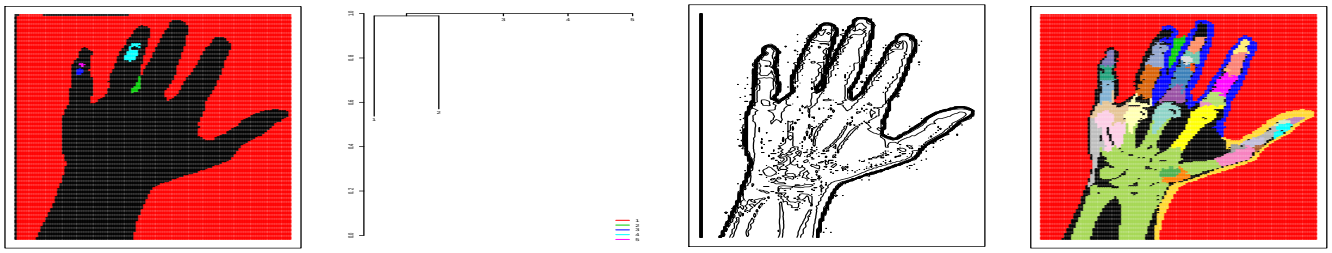

FIGURE 15 Estimated density level sets superimposed to the image (I), cluster tree (II), cluster cores of the segmentation in the top left panel of Figure 7 (III), and result of the segmentation using $h = 0.9h_N$ (IV panel).

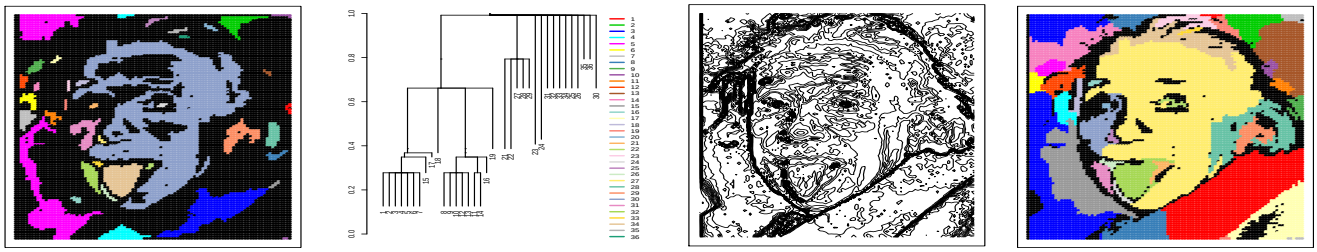

FIGURE 16 Estimated density level sets superimposed to the image (I), cluster tree (II), cluster cores of the segmentation in the top left panel of Figure 8 (III), and result of the segmentation using $h = 1.1h_N$ (IV panel).

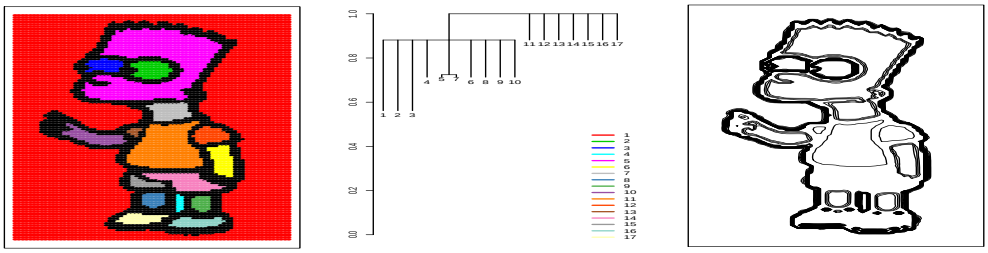

FIGURE 17 Estimated density level sets superimposed to the image (left), cluster tree (right panel) and cluster cores of the segmentation in the top left panel of Figure 9 .

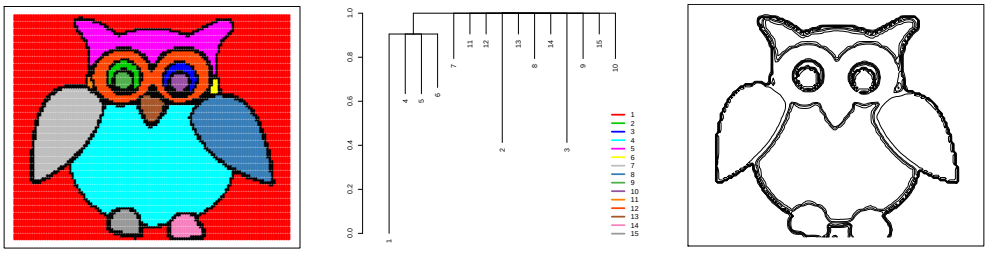

FIGURE 18 Estimated density level sets superimposed to the image (left), cluster tree (right panel) and cluster cores of the segmentation in the top left panel of Figure 10 .

K. Fukunaga and L. D. Hostetler. The estimation of the gradient of a density function, with application in pattern recognition. *IEEE Transactions on Information Theory*, 21 (1):32–40, 1975.

G. Menardi and A. Azzalini. An advancement in clustering via nonparametric density estimation. *Statistics and Computing*, 24(5):753–767, 2014.

G. Pau, F. Fuchs, O. Sklyar, M. Boutros, and W. Huber. Ebimage—an R package for image processing with applications to cellular phenotypes. *Bioinformatics*, 26(7):979–981, 2010. .

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018. URL https://www.R-project.org/.

D. Scott and S. Sain. Multidimensional density estimation. *Handbook of Statistics*, 24:229–261, 2005.

B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, New York, 1986.

P. Soille. *Morphological image analysis: principles and applications*. Springer Science & Business Media, 2013.

W. Stuetzle and R. Nugent. A generalized single linkage method for estimating the cluster tree of a density. *Journal*

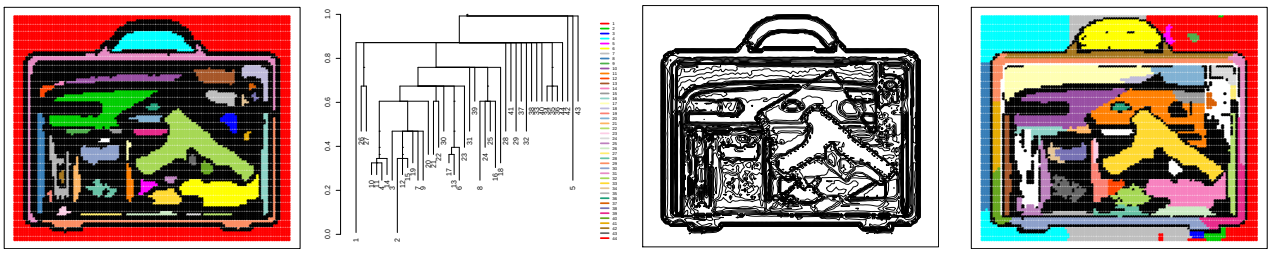

FIGURE 19 Estimated density level sets superimposed to the image (left), cluster tree (right panel) and cluster cores of the segmentation in the top left panel of Figure 11 .

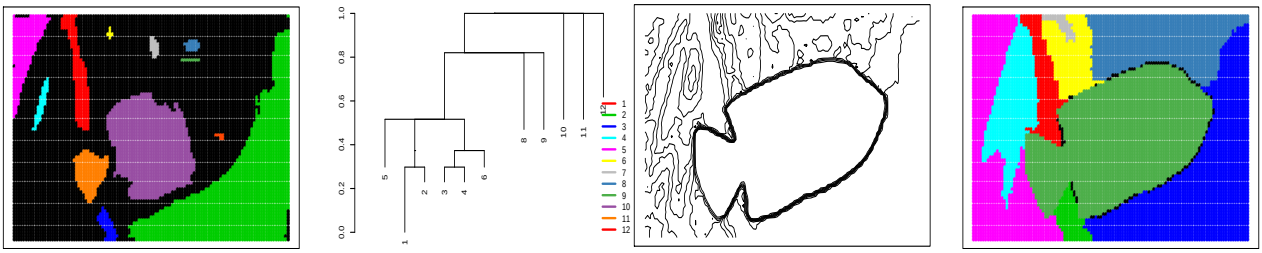

FIGURE 20 Estimated density level sets superimposed to the image (I), cluster tree (II), cluster cores of the segmentation in the top left panel of Figure 12 , and result of the segmentation using $h = 1.2h_N$

of Computational and Graphical Statistics, 19(2):397–418, 2010.

W. Tao, H. Jin, and Y. Zhang. Color image segmentation based on mean shift and normalized cuts. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 37(5): 1382–1389, 2007.

X. Yuan, B. Hu, and R. He. Agglomerative mean-shift clustering. *IEEE Transactions on Knowledge and Data Engineering*, 24(2):209–219, 2012.

How to cite this article: Menardi, G., (), Nonparametric clustering for image segmentation, , *XX*.