

Kernel-based classification using quantum mechanics

Nikolaos Nasios, Adrian G. Bors*

Department of Computer Science, University of York, York YO10 5DD, UK

Received 12 October 2005; received in revised form 26 July 2006; accepted 15 August 2006

Abstract

This paper introduces a new nonparametric estimation approach inspired from quantum mechanics. Kernel density estimation associates a function to each data sample. In classical kernel estimation theory the probability density function is calculated by summing up all the kernels. The proposed approach assumes that each data sample is associated with a quantum physics particle that has a radial activation field around it. Schrödinger differential equation is used in quantum mechanics to define locations of particles given their observed energy level. In our approach, we consider the known location of each data sample and we model their corresponding probability density function using the analogy with the quantum potential function. The kernel scale is estimated from distributions of K -nearest neighbours statistics. In order to apply the proposed algorithm to pattern classification we use the local Hessian for detecting the modes in the quantum potential hypersurface. Each mode is assimilated with a nonparametric class which is defined by means of a region growing algorithm. We apply the proposed algorithm on artificial data and for the topography segmentation from radar images of terrain.

© 2006 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.

Keywords: Kernel density estimation; Nonparametric modelling; Quantum mechanics; Vector field segmentation

1. Introduction

There are two main approaches in pattern classification: parametric and nonparametric [1,2]. The first approach assumes the existence of a statistical model for the data [1,3]. The nonparametric approach aims to approximate the probability density function (pdf) without any underlying model assumption [1,4,5]. The nonparametric methods can be classified into histogram-based and kernel-based approaches [1,5]. Histogram-based approaches require a large data set and lack the properties that would make them suitable to applications. Kernel methods result in smooth, continuous and differentiable density estimates [2,4–15].

Nonparametric approaches usually associate a function to each data sample [4,6]. Such a function is called kernel function and is characterized by a radial influence region that has its largest value at the data sample location while decreasing in value as we depart from that location. In this

case the influence of a kernel function depends on a scale parameter, called also kernel width or bandwidth. By summing up the kernel functions corresponding to a given data set we obtain an approximation of the pdf for the given data. Various nonparametric algorithms were known under different names, such as Parzen windows [4], probabilistic neural networks [8], potential functions [1], etc. Comparisons among several different kernel functions have revealed only small differences in their data modelling performance [5]. While Epanechnikov kernel is optimal when considering approximation accuracy, in most research studies the Gaussian kernel is preferred due to its simplicity, approximation capability, and differentiability properties [2,5].

Given their mathematical properties, kernel functions can be assimilated with the electro-magnetic or the gravitational field as defined in physics. In Ref. [16], the data are seen as having magnetic properties. Clusters are defined using the correlation in the orientation of Potts spins that are assigned to data. Moreover, data samples can be seen as particles that are obeying the quantum physics laws. By extending this association, a new approach was considered for the nonparametric modelling of data [17–19]. This approach relies on

* Corresponding author. Tel.: +44 1904 432780; fax: +44 1904 432767.
E-mail address: adrian.bors@cs.york.ac.uk (A.G. Bors).

the potential provided by the Schrödinger partial differential equation [20,21]. In quantum mechanics we can define the probability of locating a particle on an orbit, corresponding to a given potential energy level, by solving the Schrödinger equation. In this study we consider the reverse problem. We assume the eigenfunction (ground state) representing the lowest potential state locally, given as a sum of Gaussians, each centred at a data point and depending on a scale parameter [2]. Afterwards, the corresponding Schrödinger potential can be calculated for the given eigenfunction [18,19]. The quantum potential is thus assimilated with the pdf of the data set. By interpreting the quantum potential function we can infer information about the given data set [22].

Finding modes in nonparametric data representation has been an important task in statistics [4] and in pattern recognition [2]. While in classical kernel density estimation, the modes correspond to maxima in the resulting function [2], when interpreting the quantum potential they are represented by minima [18]. Moreover, the kernel density estimation depends on the Gaussian scale [18,19]. The location of the K -nearest neighbour has been used for calculating the scale [6]. However, this approach is sensitive to the presence of outliers in the given data set. Various studies have been undertaken for finding the appropriate kernel scale [7,9,10,15,23]. The existing methods can be classified in quality-of-fit and plug-in methods. The algorithms from the first category tend to over-peak the resulting pdf, while those from the second category oversmooth and require the knowledge of an initial scale estimate [5,7,9]. In this paper, we propose a statistical approach for estimating the kernel scale parameter. The proposed algorithm samples K -nearest neighbours [1] from the given data set and calculates their local variance [22]. A histogram of local variance is formed and a Gamma distribution is fitted to this histogram. The scale is estimated from this distribution. In Refs. [18,19], the quantum potential function minima are chosen by using the gradient descent after appropriate thresholding. However, the gradient descent algorithm is prone to getting stuck in local minima and failing to find all the modes. In this study we use the Hessian of the potential function for detecting the modes. The modes are defined according to the sign of the Hessian eigenvalues [22,25]. A region growing algorithm [26], that is a generalization of the mathematical morphology dilation operator [27] on labelled sets, is used to classify the data set into regions. The proposed approach is applied for clustering data located along two spirals circling each other which represents a difficult pattern classification problem [28,29], in blind detection of modulated signals [3], and for segmenting vector fields. The vector fields represent surface orientations extracted from a Synthetic Aperture Radar (SAR) image of terrain [30,31]. A general introduction in nonparametric modelling is provided in Section 2, while the proposed nonparametric approach derived from quantum mechanics is described in Section 3. The estimation of the scale parameter is explained in Section 4, and the algorithm for identifying the modes is detailed in Section 5. The experimental results are provided

in Section 6 and the conclusions of this study are drawn in Section 7.

2. The nonparametric methodology

When no model assumptions can be formulated about the given data set we should use a nonparametric approach [1,5]. While data modelling, using “naive” nonparametric estimators such as histograms lacks interpretability and generalization capability, kernel-based estimators provide smooth functions that asymptotically represent the true pdf. Kernel-based estimation has received an increased attention in statistics, pattern recognition and computer vision research [2,7,10,12–15]. A kernel function is assigned to each data sample \mathbf{X}_i , $i = 1, \dots, N$ and the pdf is approximated by

$$\psi(\mathbf{X}) = \frac{1}{N\sigma^d} \sum_{i=1}^N \mathcal{K} \left(\frac{\mathbf{X} - \mathbf{X}_i}{\sigma} \right), \quad (1)$$

where $\mathcal{K}(\cdot)$ is the kernel function which is assumed homoscedastic, σ is a parameter called scale, width or bandwidth, and d is the data dimension. The kernel with optimal efficiency was found to be the Epanechnikov kernel [5,13]:

$$\mathcal{K}_E(\mathbf{X}_i) = \begin{cases} \frac{d+2}{2C_d} [1 - (\mathbf{X} - \mathbf{X}_i)^T \\ \times (\mathbf{X} - \mathbf{X}_i)] & \text{if } (\mathbf{X} - \mathbf{X}_i)^T (\mathbf{X} - \mathbf{X}_i) < 1, \\ 0 & \text{otherwise,} \end{cases} \quad (2)$$

where \mathbf{X}_i are d -dimensional data samples $i = 1, \dots, N$, C_d is the volume of the d -dimensional sphere whose radius is one. However, due to its smoothness and differentiability properties, the preferred kernel is the Gaussian kernel in parametric [3] and nonparametric approaches [2,5,7,10,18]:

$$\mathcal{K}_G(\mathbf{z}_i) = \exp \left(-\frac{\mathbf{z}_i^T \mathbf{z}_i}{2} \right), \quad (3)$$

where

$$\mathbf{z}_i = \frac{\mathbf{X} - \mathbf{X}_i}{\sigma} \quad (4)$$

is a scaled variable. The function (1) that uses the Gaussian kernel is also called scale-space function.

The pdf approximation by Eq. (1) results in a smooth function that represents well the data. One important task in nonparametric classification is to interpret the resulting function (1) [4,6]. However, the approximation of the given data pdf depends on the choice of the parameter σ [5,7,23]. The mean shift algorithm relies on the estimation of the gradient of the density function $\psi(\mathbf{X})$ from Eq. (1), instead of the function itself [11–15]. The mean shift updating when using the Epanechnikov kernel (2) [13,15] is given by

$$M_\sigma(\hat{\mathbf{C}}_{t+1}) = \frac{\sigma^2}{d+2} \frac{\nabla \psi(\hat{\mathbf{C}}_t)}{\psi(\hat{\mathbf{C}}_t)} = \frac{1}{N_\sigma} \sum_{\mathbf{X}_i \in S_\sigma(\hat{\mathbf{C}}_t)} \mathbf{X}_i - \hat{\mathbf{C}}_t, \quad (5)$$

where N_σ is the cardinality of the samples located in the hypersphere $S_\sigma(\hat{\mathbf{C}}_t)$, of a radius equal to σ and $\hat{\mathbf{C}}_{t+1}$ is the updated cluster centre at iteration $t + 1$, considered initially as one of the data samples. As it can be observed from Eq. (5), the mean shift vector $M_\sigma(\hat{\mathbf{C}}_{t+1})$ points towards the direction of the steepest slope of the density function $\psi(\mathbf{X})$ leading to the modes of the underlying density [13]. The mean shifts depend on the scale σ , that defines the influence area $S_\sigma(\hat{\mathbf{C}}_t)$, and determines the cardinality N_σ as well. After an iterative process, the mean shift leads to the pdf modes. If several mode candidates are within the scale range, mode splitting or merging is decided. Two different modes are decided whenever there is a local minima on the line that joins them [13,14].

3. Quantum clustering

The kernel function can be assimilated with the function representing the energy field around a particle in physics. In this situation, the effect of several particles, associated with data samples, results into a potential energy function, such as the one described by Eq. (1). In the following, we describe a new approach which is inspired from quantum physics.

3.1. Quantum physics foundation

Quantum physics emerged from the failures of classical mechanics. The uncertainty principle of Heisenberg states that the position and the momentum of a particle cannot be determined simultaneously. The six quantum mechanics postulates along with Heisenberg's uncertainty principle define the mathematical foundation for key aspects of quantum mechanics [20,21].

The intuition behind a clustering approach based on quantum physics principles relies on the analogy between data points and electrons or other particles in a certain state. The state of a quantum mechanics system is completely specified by a wavefunction $\psi(\mathbf{X}, t)$ that vanishes at infinity. According to the first postulate of quantum mechanics, the probability that a particle lies in a volume element $d\mathbf{X}$, located at \mathbf{X} , at time t , is given by [20]:

$$P(\mathbf{X}, t) = \int |\psi(\mathbf{X}, t)|^2 d\mathbf{X}. \quad (6)$$

This equation implies that with the time passing, it is less likely to find the particle at its initial location. The wavefunction $\psi(\mathbf{X}, t)$ can be used to calculate various physical quantities relevant to the given quantum system [21].

Similarly to a quantum system formed by particles, the statistical interdependence among data points can be described with the Schrödinger differential equation, after imposing some weak constraints. Such an analogy, can be interpreted as representing the data points by a wavefunction $\psi(\mathbf{X}, t)$ at location \mathbf{X} and time t , as in the case of particles in a quantum mechanics system. The wavefunction

plays the role of an activation field, similar with that corresponding to the electro-magnetic field of a particle, around each data point.

According to the fifth postulate of quantum mechanics, a quantum system evolves in space and time following the Schrödinger differential equation [20]. The Schrödinger equation predicts the behaviour of a quantum mechanics system and is completely specified by a wavefunction $\psi(\mathbf{X}, t)$, which estimates analytically the location probability for a particle, in space and time, subject to a degree of uncertainty. In the following, we drop the variation with respect to time and consider only the dependency of the quantum system with respect to its location as $\psi(\mathbf{X})$. The time-independent Schrödinger equation is given by

$$\mathcal{H}\psi(\mathbf{X}) \equiv \left(-\frac{\sigma^2}{2} \nabla^2 + V(\mathbf{X}) \right) \psi(\mathbf{X}) = E_g \cdot \psi(\mathbf{X}), \quad (7)$$

where \mathcal{H} is the Hamiltonian operator, E_g is an eigenvalue energy level, $\psi(\mathbf{X})$ corresponds to the eigenstate of the given quantum system, $V(\mathbf{X})$ is the Schrödinger potential and ∇^2 is the Laplacian. The potential is always positive, $V(\mathbf{X}) \geq 0$. This equation constitutes a basic postulate in quantum physics theory which has been confirmed by innumerable experiments assessing that the calculated level of energies E_g correspond to the measured values [21]. The wavefunction $\psi(\mathbf{X})$ is a solution of equation (7) and describes the evolution of physical processes associated with the given quantum mechanics system. Conventionally, the potential $V(\mathbf{X})$ is given and the equation is solved to find solutions $\psi(\mathbf{X})$. Eq. (7) is used by physicists to calculate orbits of electrons and other particles in atoms for a specific energy level $V(\mathbf{X})$, that can be either measured experimentally or predicted theoretically [21]. Such a function $\psi(\mathbf{X})$, can be assimilated with the kernel-based sum from Eq. (1), that depends on data coordinates \mathbf{X}_i .

3.2. Data representation using quantum potential

Unlike quantum physics, where we want to estimate the location of particles given their energy levels, in nonparametrical clustering we consider the inverse problem. In this case, we assume known the location of data samples and their state as given by Eq. (1) which is considered as a solution for Eq. (7), subject to a set of constraints. The algorithm developed on these principles is called quantum clustering [17–19]. By viewing the wavefunction $\psi(\mathbf{X})$ as an eigenstate of the Schrödinger equation, the goal in quantum clustering becomes to estimate the quantum potential $V(\mathbf{X})$, which describes a hypersurface and characterizes the data probability density function. Similarly to particles of equal potential that are characterized by a certain state, data points that are located nearby have close potential values. In this case, we can observe the analogy between clusters of data points and quantum particles that are characterized by a similar state and which revolve in neighbouring orbits. Another analogy

is between particles, characterized by low potential values, and the minima of the data potential function specific to cluster centres.

In a pattern classification application we want to evaluate the potential $V(\mathbf{X})$ produced by the quantum mechanical system which is assimilated with the given data. We can observe that the solution from Eq. (7), for a single data point, $N = 1$, \mathbf{X}_1 , for the system given when considering Eq. (3) as a kernel, is a quadratic expression:

$$V(\mathbf{X}) = \frac{(\mathbf{X} - \mathbf{X}_1)^T(\mathbf{X} - \mathbf{X}_1)}{2\sigma^2} \quad (8)$$

and this potential corresponds to the ground state of the quantum mechanical system defined by the given data point [20,21].

After replacing $\psi(\mathbf{X})$ from Eqs. (3), (1) into Eq. (7), we solve the Schrödinger potential for the given data set as [18,19]:

$$V(\mathbf{X}) = E - \frac{d}{2} + \frac{1}{2\sigma^2\psi(\mathbf{X})} \sum_{i=1}^N (\mathbf{X} - \mathbf{X}_i)^T(\mathbf{X} - \mathbf{X}_i) \times \exp \left[-\frac{(\mathbf{X} - \mathbf{X}_i)^T(\mathbf{X} - \mathbf{X}_i)}{2\sigma^2} \right], \quad (9)$$

where E is a constant corresponding to the energy eigenvalue, such that we enforce $V(\mathbf{X}) \geq 0$. This equation describes the potential hypersurface corresponding to the given data set and can be interpreted as a cost function that is used to cluster data. The method that uses (9) for associating data into classes is called the quantum clustering algorithm. The potential hypersurface corresponds to the pdf for the given data set. The denominator from the second part of Eq. (9) produces local minima for $V(\mathbf{X})$ in the areas with high data concentration, i.e. large $\psi(\mathbf{X})$. On the other hand the Laplacian tends to spread the potential surface away from the potential function minima. Both effects depend on the non-linear weighting with the scale parameter σ . The modes of the potential function (9) are associated with data classes or clusters. The potential function modes are local minima in the expression represented by Eq. (9) [19,22]. Just as particles of low potential value are stabilized, since they minimize their vibration, the minima of the potential function can be associated with cluster centres.

4. Kernel scale estimation

In general, the degree of the kernel function influence in the potential function is controlled by a scale parameter σ as it can be observed from Eq. (1). The potential function can be smooth or over-peaked if σ is either larger or smaller [18,23]. Implicitly, the number of classes in the data depends on the appropriate selection of σ . Quite often, σ is initialized to arbitrary values [18,19], or by employing certain test hypotheses [7,9,15,23].

In this paper we propose a new statistical approach for estimating the scale σ . A straight forward consideration is to assume that the scale parameter σ is associated with the local data spread. Under this assumption, $\hat{\sigma}$ can be calculated from the statistics of Euclidean distances between samples from localized data sets. Defining the scale just by the distance to the K th-nearest neighbour as in Ref. [6], is prone to the influence by noise and outliers. On the other hand, K -nearest neighbours have been successfully used in pattern recognition for taking into account the local data distribution in the estimation process [1]. In our approach, we consider the local distribution of data and we estimate the variance of K -nearest neighbours to a given data sample \mathbf{X}_i . Let us randomly sample a data \mathbf{X}_i from our data set. Considering this data sample, we rank all the other data according to their Euclidean distance to \mathbf{X}_i :

$$\{\mathbf{X}_{(k)}, k = 1, \dots, K \mid \|\mathbf{X}_{(k)} - \mathbf{X}_i\|^2 < \|\mathbf{X}_{(k+1)} - \mathbf{X}_i\|^2\}, \quad (10)$$

where $\mathbf{X}_{(k)}$ represent the K -nearest neighbours of \mathbf{X}_i , sampled from the entire data set with $\mathbf{X}_{(1)}$ the closest data sample to \mathbf{X}_i , while $\mathbf{X}_{(K+1)}$ is outside the chosen neighbourhood data set. Let us assume that we repeat this operation for M data samples, $i = 1, \dots, M$, $M \leq N$. We calculate the variance corresponding to these localized neighbourhoods:

$$s_i = \frac{\sum_{k=1}^K \|\mathbf{X}_{(k)} - \mathbf{X}_i\|^2}{K - 1}, \quad (11)$$

where $K < N$ is the cardinality of a data set that defines the chosen neighbourhood. The histogram of s_i is defined by calculating the local variance (11) of K -nearest neighbours around a chosen data subset \mathbf{X}_i , $i = 1, \dots, M$.

The variance distribution for data samples that are generated by normal independent distributions with mean 0 and variance 1 are modelled by a chi-square distribution [5]. Gamma distribution is a generalization of the chi-square distribution and is suitable for modelling distributions of data sample variance in the case when we have no knowledge about the underlying data distribution [22]. The Gamma distribution depends on two parameters [24]:

$$p(s) = \frac{s^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} e^{-s/\beta}, \quad (12)$$

where $\alpha > 0$ is the shape parameter that enables the function to have a variety of shapes, $\beta > 0$ is the scale parameter of the Gamma distribution. When β is larger or smaller than 1, it allows the density function to stretch or compress, respectively. $\Gamma(\cdot)$ represents the Gamma function:

$$\Gamma(t) = \int_0^\infty r^{t-1} e^{-r} dr. \quad (13)$$

By modelling distributions of variance of K -nearest neighbours we have a statistical description of the second-order moment in localized data sets.

The parameters α and β are estimated from the empirical distribution of localized data variance calculated by Eqs. (10) and (11). A method to calculate the parameters of the Gamma distribution is the moments method [24]. We evaluate the first and second order moments of the given distribution. The mean of the empirical distribution is given by

$$\bar{s} = \frac{\sum_{i=1}^M s_i}{M}, \tag{14}$$

where s_i is estimated in Eq. (11), $i = 1, \dots, M$ and M is the number of sampled local neighbourhoods. The variance of the data histogram corresponding to the distribution of local variance is calculated as

$$l = \frac{\sum_{i=1}^M (s_i - \bar{s})^2}{M - 1}. \tag{15}$$

Given the two moments, the Gamma distribution parameters are estimated as [24]:

$$\hat{\alpha} = \left(\frac{\bar{s}}{l}\right)^2, \tag{16}$$

$$\hat{\beta} = \frac{l^2}{\bar{s}}, \tag{17}$$

where \bar{s} and l are evaluated in Eqs. (14) and (15), respectively. We consider the Gamma distribution, resulted from the fitting to the histogram of local variance, as modelling the statistics of the scale estimate $\hat{\sigma}$.

5. Classification using the potential function modes

Finding modes in a nonparametric pdf representation is considered an important step in data analysis [2,4,5]. Each mode can be associated with a class of data samples [12,19]. After finding an appropriate scale estimate we represent the data pdf by means of the scale-space function $\psi(\mathbf{X})$ from Eq. (1), or the quantum potential function $V(\mathbf{X})$ resulting from applying the Schrödinger equation (9). Such a function can be interpreted as a landscape in the $(d + 1)$ th dimension. The quantum potential depends on the second derivative of the data pdf. Consequently, in the case of $\psi(\mathbf{X})$ the modes are represented by local maxima, while in the case of $V(\mathbf{X})$ they are the local minima. Let us assume a regular orthogonal lattice \mathbf{Z} , that is defined by sampling at regular intervals within data range along each dimension, $k = 1, \dots, d$:

$$\mathbf{Z} = \{(Z_{1,i}, Z_{2,i}, \dots, Z_{d,i})^T \mid \|Z_{k,i} - Z_{k,i-1}\| = D_k\}, \tag{18}$$

where \mathbf{Z}_i is a specific knot on the lattice, which is bounded by the extreme data sample values on k th dimension, and D_k is the inter-knot distance on the lattice along k th axis. Either $\psi(\mathbf{Z})$ or $V(\mathbf{Z})$ is evaluated for the knots on this lattice by considering either functions (1) or (9), respectively. The inter-knot distance is constant along each direction and

depends on the kernel scale:

$$D_k = \frac{\hat{\sigma}^2}{2} \tag{19}$$

for $k = 1, \dots, d$, where $\hat{\sigma}$ is the scale estimate drawn from the Gamma distribution (12). Replacing the evaluation of the pdf of a given data set with its calculation on a fixed lattice results in a smooth functional representation while the required computational complexity can be easily assessed.

5.1. Interpreting the local Hessian

The Hessian has been used for finding local extremes in functions for various applications [25]. If we assume $d = 2$, the Hessian is calculated on the given lattice as:

$$\mathbf{H}[F(\mathbf{Z})] = \left(\frac{\partial^2 F(\mathbf{Z})}{\partial x \partial y} \right)_{x,y}, \tag{20}$$

where $F(\mathbf{Z})$ can be either $\psi(\mathbf{Z})$ from Eq. (1) or $V(\mathbf{Z})$ from Eq. (9). The evaluation of a potential function on a regular lattice as that described by Eq. (18) facilitates the calculation of the local Hessian by considering simple differences of the function values at neighbouring lattice knots on each data dimension. The eigendecomposition of the Hessian matrix provides:

$$\mathbf{H} = \mathbf{T} \cdot \mathbf{\Lambda} \cdot \mathbf{T}^{-1}, \tag{21}$$

where $|\cdot|$ denotes matrix multiplication, \mathbf{T} is a matrix whose columns represent the eigenvectors, while $\mathbf{\Lambda} = \{\lambda_k \mid k = 1, \dots, d\}$ is a diagonal matrix that contains the eigenvalues λ_k . The sign of the eigenvalues of the Hessian matrix are used to identify the local extremes associated with the modes [25]. A saddle point occurs where there is simultaneously at least one local maximum and one local minimum in two directions that are orthogonal to each other. Following the eigendecomposition, we can identify the local minima, maxima and saddle points according to the signs of the eigenvalues for the entire lattice:

$$\begin{aligned} \lambda_k(\mathbf{Z}) > 0, \quad \forall k = 1, \dots, d \text{ then local minimum,} \\ \lambda_k(\mathbf{Z}) < 0, \quad \forall k = 1, \dots, d \text{ then local maximum,} \\ \exists \lambda_j(\mathbf{Z}) > 0, \wedge \exists \lambda_i(\mathbf{Z}) < 0, \quad i \neq j \text{ then saddle point.} \end{aligned} \tag{22}$$

A common sense assumption is that either local minima or local maxima are surrounded by saddle points [14]. This information is used in the following to define the modes.

5.2. Nonparametric classification using the potential function

A class contains all the data that lie in a compact area of the lattice where all the lattice knots correspond to local maxima, when the kernel density is defined as in Eq. (1), and are surrounded by local minima and saddle points. Conversely, when using Eq. (9) a class is made up of all the data from a compact area that consist of lattice knots whose

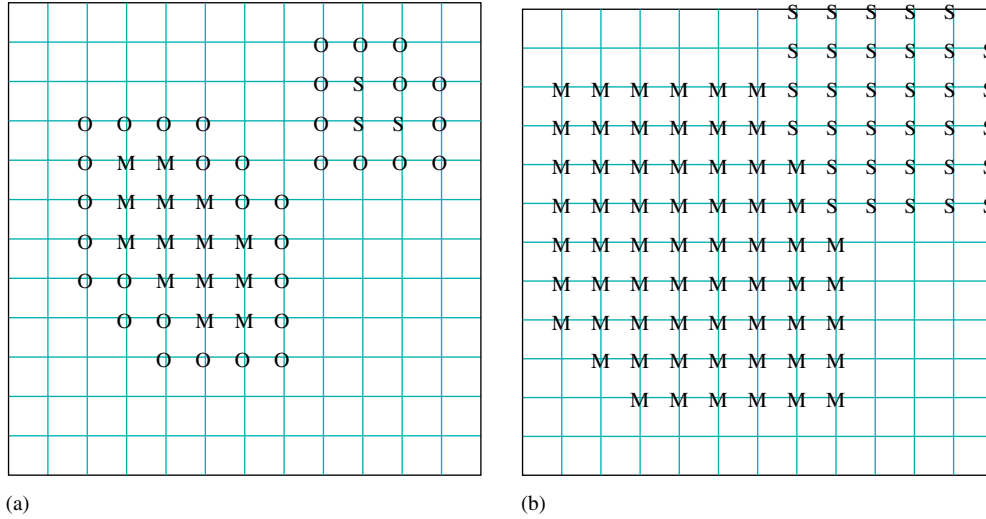


Fig. 1. Defining the modes of a potential function on a lattice: (a) modes of the potential; (b) splitting into clusters.

potential fulfil the condition of local minima and are surrounded by local maxima and saddle points. Following these assumptions, we can proceed to split the lattice in various regions according to their corresponding modes. In order to define such regions we use eight-site neighbourhoods denoted as $\mathcal{N}(\mathbf{Z}_i)$, where \mathbf{Z}_i is the lattice knot. We check for connectivities among lattice knots that have identical signs for their eigenvalues of the potential function Hessian. The resulting regions, when the scale-space function is modelled by Eq. (1), correspond to local clusters defined as

$$\{\mathbf{Z}_j \mid \lambda_k(\mathbf{Z}_j) < 0, \forall k = 1, \dots, d, \mathbf{Z}_j \in \mathcal{N}(\mathbf{Z}_i)\} \\ \text{while } \mathbf{Z}_j \notin \mathcal{N}(\mathbf{Z}_l), l \neq j\}, \quad (23)$$

where j and l are two different local modes that are separated by saddle points and have no connectivity to each other. In Fig. 1(a) the procedure of defining the clusters is shown. With “M” and “S” we represent two local maxima when we have Eq. (1) or two local minima if the potential is given by Eq. (9). Each class is defined by a mode \mathbf{Z}_j in Fig. 1, while their neighbourhoods $\mathcal{N}(\mathbf{Z}_j)$ are marked with “O” and correspond to saddle points and local functional extremes that are opposite to those of the modes.

After defining the modes of the potential function we evaluate their significance in the pdf representation of the given data set. A factor L_k is assigned to each mode, calculated as the ratio of its energy from the energy of all the modes defined in the given potential function:

$$L_i = \frac{\int_{\mathbf{Z}_i} F(\mathbf{Z}) d\mathbf{Z}}{\sum_k \int_{\mathbf{Z}_k} F(\mathbf{Z}) d\mathbf{Z}}, \quad (24)$$

where $F(\mathbf{Z})$ is the potential function defined as $\psi(\mathbf{Z})$ from Eq. (1) or as $V(\mathbf{Z})$ from Eq. (9). The calculation of the potential function onto the lattice defined by Eq. (18) results in evaluating the integrals as simple Riemann additions.

The modes are afterwards ordered according to their decreasing significance from the total potential.

After associating each mode to a class we can proceed to split the given data set among classes. The following procedure does not require any parametric knowledge. The region labelling process continues with a region growing algorithm that works on labelled regions [26] and consists of the generalization of the binary dilation operator, known in mathematical morphology [27], under a set of local conditions. At each iteration, a layer of lattice knots are added simultaneously to each cluster as represented by the points marked with “O” in Fig. 1(a). This process continues until lattice regions corresponding to two different clusters become adjoint and ends when all the knots on the lattice are assigned to one of the classes. This algorithm preserves the shape of the classes without considering any parameters, and corresponds to an unsupervised nonparametric classification approach, by associating lattice knots with a specific mode. After labelling all the lattice knots according to the algorithm, as illustrated in Fig. 1(b), we assign data samples to one or another of the clusters according to their minimal distance to the labelled knots:

$$\left\{ \mathbf{X}_i \in \mathcal{C}_k \mid k = \arg \min_j \|\mathbf{X}_i - \mathbf{Z}_j\|^2 \right\}, \quad (25)$$

where \mathcal{C}_k denotes the class associated with mode k , and \mathbf{Z}_j are lattice knots assigned to class $j = 1, \dots, G$, where G is the total number of detected modes.

6. Experimental results

The proposed methodology is applied to various artificial and real data sets. In the first example, we use the proposed clustering methodology in order to separate two spiral patterns. This data set represents a difficult problem

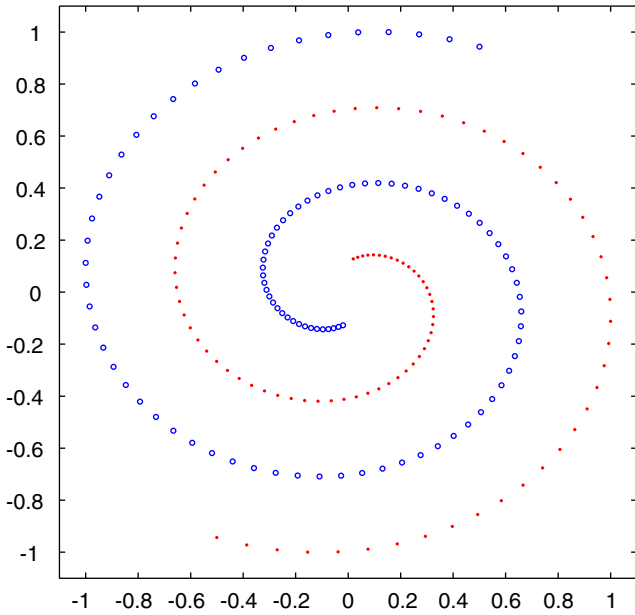


Fig. 2. Two spiral classes dataset.

that has been considered as a benchmark in pattern classification [28,29]. The data corresponding to the two clusters are generated in polar coordinates according to

$$\theta_{1,i} = -\frac{2\pi}{3} + i \frac{\pi}{32}, \quad (26)$$

$$\theta_{2,i} = \frac{\pi}{3} + i \frac{\pi}{32}, \quad (27)$$

$$R_{1,i} = R_{2,i} = R_{\max} \frac{114 - i}{114} \quad (28)$$

for $i = 1, \dots, N$, where $\{\theta_{1,i}, R_{1,i}\}$ represents the polar coordinates of a data sample i from the first cluster and $\{\theta_{2,i}, R_{2,i}\}$ are the polar coordinates of a data sample from the second cluster. We assume $R_{\max} = 1$ and $N = 100$. The coordinates of one cluster are $\mathbf{X}_1 = (R_1 \cos \theta_1, R_1 \sin \theta_1)$ while for the second cluster they are $\mathbf{X}_2 = (R_2 \cos \theta_2, R_2 \sin \theta_2)$. The two spiral classes are represented in Fig. 2. After applying the clustering methodology described in the paper we evaluate the function $\psi(\mathbf{Z})$ defined in Eq. (1) and its result on the two spiral data is shown in Fig. 3(a). The quantum potential $V(\mathbf{Z})$ calculated according to Eq. (9) is evaluated on the same data in Fig. 3(b). From Fig. 3 we can observe that both methods appropriately detected two clusters. However, the minimas between the two spirals are better defined by the quantum clustering, while the scale-space method, defined by $\psi(\mathbf{Z})$, provides a large maxima in the origin of the spirals and a significant gradient along each of the spiral which may be misleading.

Nonparametric classification algorithms using the clustering methodology described in this paper are applied to two artificial data sets for blind detection of modulated signals. We consider two cases of modulated signals: quadrature amplitude modulated (QAM) signals and phase-shifting-key

(PSK) modulated signals. These data have been used for assessing a variational Bayesian approach in Ref. [3]. The perturbation channel equations when assuming inter-symbol and co-channel interference, considered in the case of 8-PSK signals are provided by

$$\begin{aligned} x_I(t) &= I(t) + 0.2I(t-1) - 0.2Q(t) \\ &\quad - 0.04Q(t-1) + N(0, 0.11), \\ x_Q(t) &= Q(t) + 0.2Q(t-1) + 0.2I(t) \\ &\quad + 0.04I(t-1) + N(0, 0.11), \end{aligned} \quad (29)$$

where $(x_I(t), x_Q(t))$ represent the in-phase and quadrature signal components at time t on the communication line, and $I(t)$ and $Q(t)$ correspond to the signal symbols (there are eight signal symbols in 8-PSK, equidistantly located on a circle of centre zero). To these signals we also consider additive Gaussian noise that corresponds to SNR = 22 dB. We have generated $N = 960$ signals, by assuming equal probabilities for all inter-symbol combinations. For 4-QAM signals we consider only additive noise, with SNR of 8 dB. The resulting signal constellations are displayed in Figs. 4(a)–(b), where each signal is represented by a single dot. It can be observed from these figures that the signal clusters corresponding to the data sets are overlapping each other. The aim of the blind detection application is to find the class corresponding to each signal despite the mixing statistics.

We sample neighbourhoods of $K = N/4$ data samples. We calculate the variance of these K -nearest neighbours and consequently form empirical distributions. According to the description from Section 4 we fit a Gamma distribution to the distribution of local variance. The distribution of local variance calculated for 4-QAM and 8-PSK are shown in Figs. 5(a) and (b), respectively. We can see a good fit between the estimated Gamma distribution and the empirical histograms. After fitting the Gamma distribution, as described in Section 4, we consider a homoscedastic kernel of scale $\hat{\sigma}$, equal to the Gamma distribution mean:

$$\hat{\sigma} = \hat{\alpha}\hat{\beta}, \quad (30)$$

where $\hat{\alpha}$ and $\hat{\beta}$ are evaluated in Eqs. (16), (17). This value can be also obtained directly by averaging s_i , using Eq. (14). The Gamma distribution and its mean are shown in the plots from Fig. 5. Using $\hat{\sigma}$ we evaluate the function $\psi(\mathbf{Z})$ defined in Eq. (1), and the quantum potential $V(\mathbf{Z})$ according to Eq. (9), on a lattice whose knots are equidistant, as provided by Eqs. (18), (19). In Figs. 6(a)–(b) the two functions, $\psi(\mathbf{Z})$ and $V(\mathbf{Z})$, are shown for 4-QAM, while in Figs. 6(c) and (d) they are calculated for 8-PSK data, respectively. The knots assigned to the modes according to the signs of their local Hessian eigenvalues (22), as described in Section 5, calculated when using the quantum potential algorithm, are marked with “*” in the plots from Fig. 4. In Fig. 6 the knots marked with “*” indicate the modes decided for both potential functions $\psi(\mathbf{Z})$ and $V(\mathbf{Z})$, respectively.

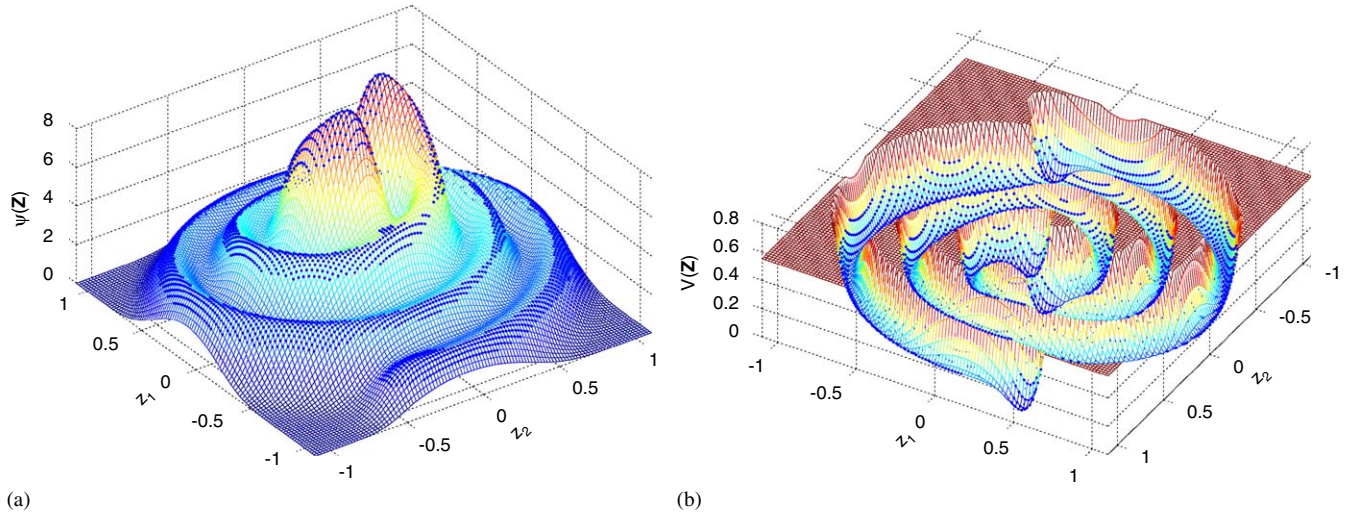


Fig. 3. The separation of two spirals using the potential functions: (a) $\psi(\mathbf{Z})$; (b) quantum potential $V(\mathbf{Z})$.

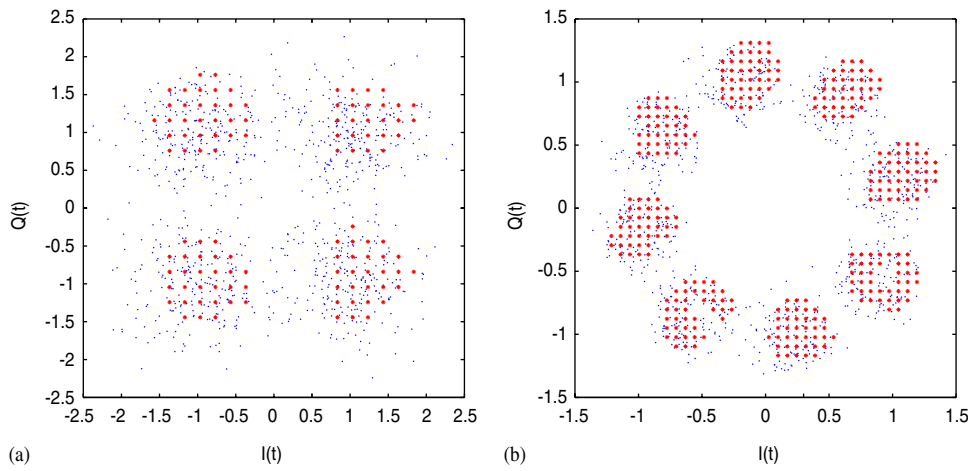


Fig. 4. Constellations of modulated signal samples, denoted by “.”, where knots associated with the detected modes are marked by “*”: (a) 4-QAM; (b) 8-PSK.

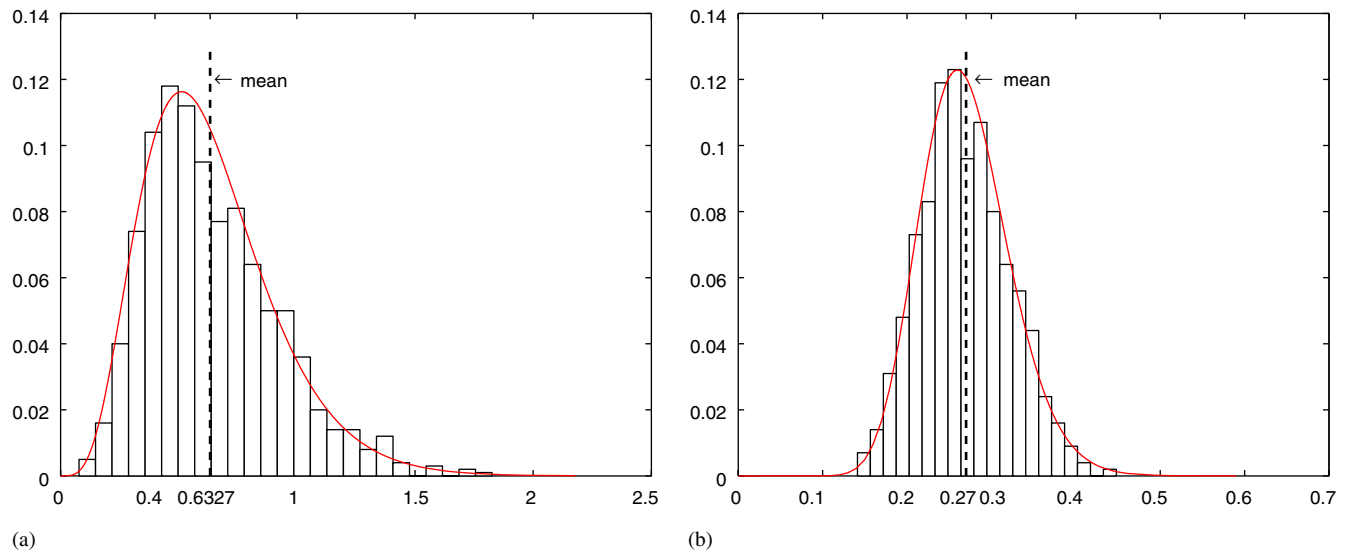


Fig. 5. Evaluating the scaling parameter: (a) 4-QAM; (b) 8-PSK.

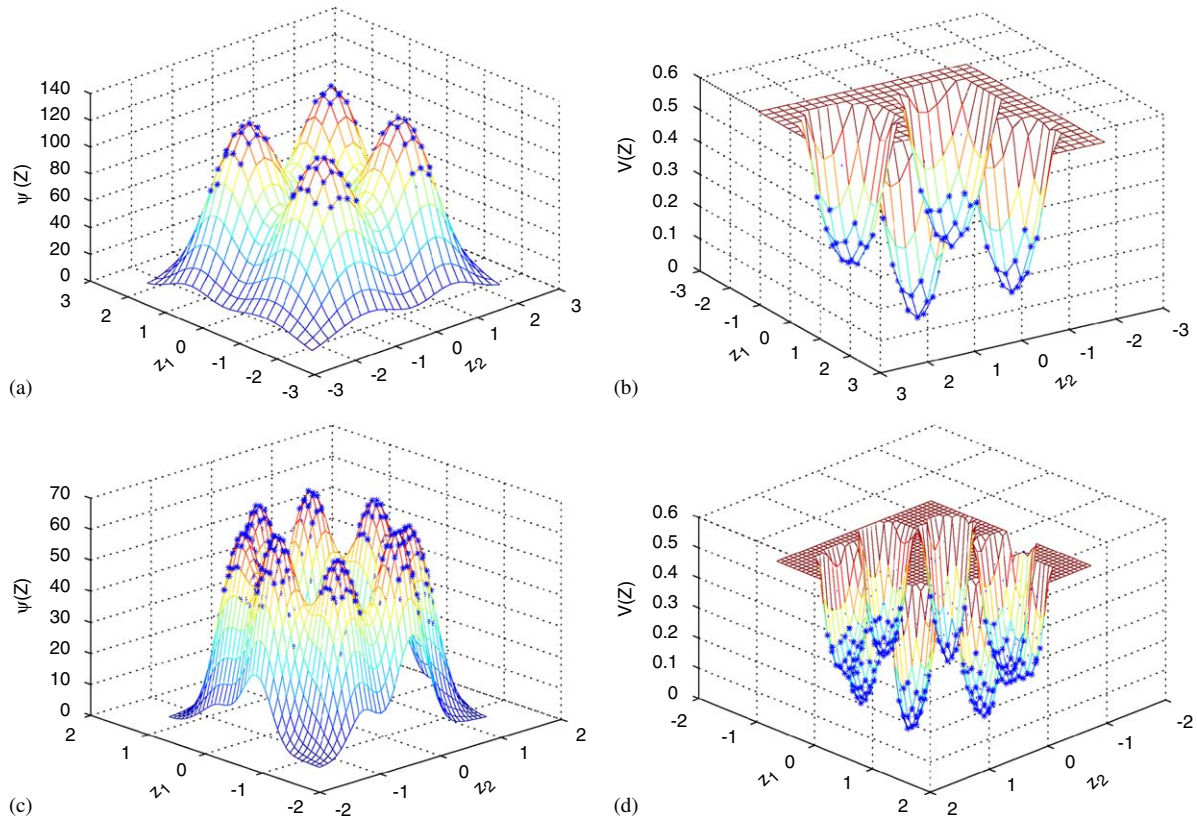


Fig. 6. Functional representations for the given data. The representation for $V(\mathbf{Z})$ in (b) and (d) is visualized from below in order to evidenciate its modality: (a) $\psi(\mathbf{Z})$ for 4-QAM; (b) $V(\mathbf{Z})$ for 4-QAM; (c) $\psi(\mathbf{Z})$ for 8-PSK; (d) $V(\mathbf{Z})$ for 8-PSK.

From Fig. 6 it can be observed that 4 modes are found in the case of 4-QAM while 8 modes are detected in the case of 8-PSK, when evaluating either $\psi(\mathbf{Z})$ or $V(\mathbf{Z})$. Consequently, the potential surface is split in regions according to the signs of the local Hessian eigenvalues (22).

The number of clusters has been determined according to the approach described in Section 5 when varying the scale σ . The plots representing the number of detected modes are shown in Fig. 7 for various σ , for both 4-QAM and 8-PSK data when considering $\psi(\mathbf{Z})$ and $V(\mathbf{Z})$, respectively. We calculate the misclassification error for the functions that correspond to valid values of σ , i.e. for the range that provides 4 modes for 4-QAM and 8 for 8-PSK, respectively. Each data sample is assigned to a class according to the potential function partition described in Section 5. This assignment is compared with the true label of that data sample, that is known when generating the artificial data set, for example according to Eq. (29) for 8-PSK. In Fig. 8 we display the misclassification error as a function of scale σ . We can observe from these plots that the estimated $\hat{\sigma}$ provides a low, close to optimal, misclassification rate for both algorithms in both data sets. Due to the overlap among the signal classes the optimal misclassification error is different from zero.

In the second set of experiments we have considered the segmentation of topography. The proposed segmentation

approach uses exclusively the classification of vector normals representing surface normals orientation. The vector field of surface normals were extracted from Synthetic Aperture Radar (SAR) images of terrain using shape from shading algorithms adapted to the corresponding illuminance model and image statistics as described in Refs. [30,31]. An area from Wales is shown in the SAR image from Fig. 9(a), while an image of Titan, one of the moons of Saturn, sent to Earth in November 2004 by the space station Cassini–Huygens, is shown in Fig. 9(c). Their corresponding surface normals after being smoothed by the vector median algorithm [32] are displayed in Figs. 9(b)–(d), respectively. These data are represented using the surface normal vector entries along x - and y -axis and are displayed in Fig. 10 for the two data sets, where the value in the centre, of coordinates (0, 0), is replicated several times. All data are contained inside a circle of radius one due to the normalization constraint imposed on surface normals. The localization of some of these data on circular arcs is due to the way how they were processed from the SAR images [30]. It can be observed from Fig. 10 that surface normal data are characterized by non-Gaussian statistics. In this case, a nonparametric method is the most appropriate approach to be used for segmenting the local topography, when using the surface normal orientations as feature vectors.

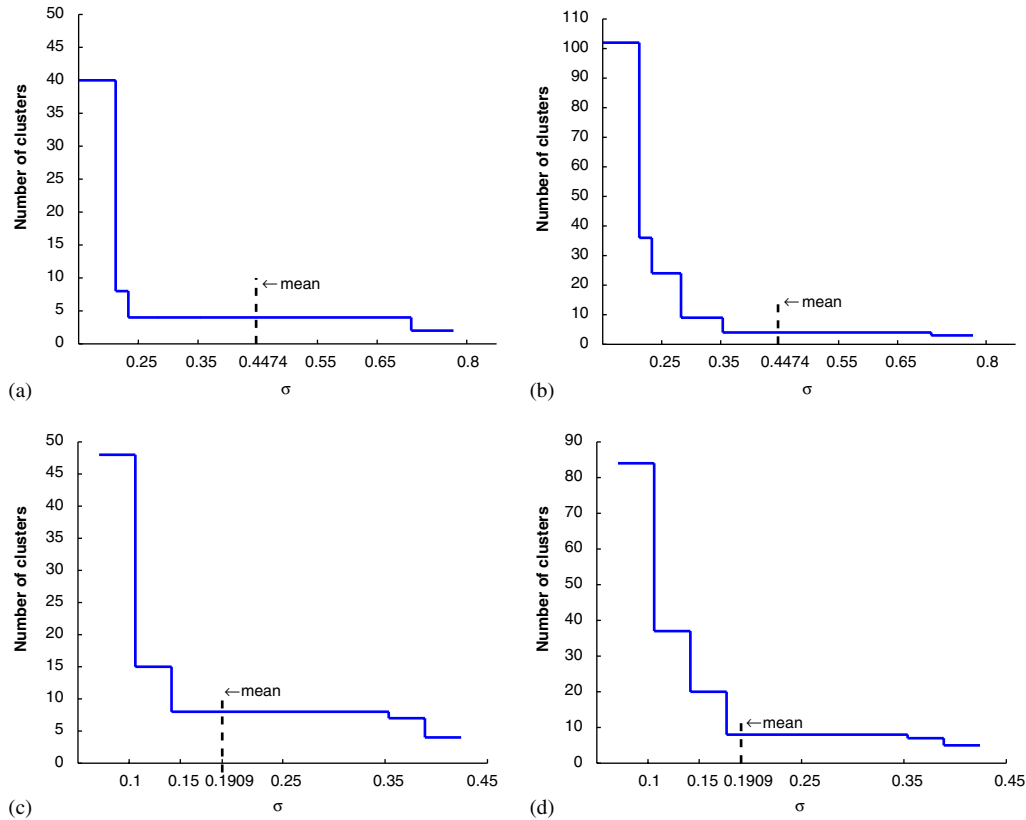


Fig. 7. Number of modes for each nonparametric method and for each data set: (a) $\psi(\mathbf{X})$ for 4-QAM; (b) $V(\mathbf{X})$ for 4-QAM; (c) $\psi(\mathbf{X})$ for 8-PSK; (d) $V(\mathbf{X})$ for 8-PSK.

For these data we consider K , the number of nearest neighbours, as a random variable with a uniform distribution, in the range $[N/20, N/2]$, where N is the number of surface normal vectors. We sample sets of each K data and we calculate their variance. The histograms of variance formed for the two data sets are displayed in Fig. 11(a) for the Wales SAR image and in Fig. 11(b) for the Titan data. We model the resulting histograms of variance with Gamma distributions. We can see from Fig. 11 a very good fit of the Gamma distribution with the actual data histograms. Three nonparametric approaches have been applied to these data sets. The first one, called the scale-space algorithm, applies the methodology presented in Section 5 to the surface represented by $\psi(\mathbf{Z})$, defined in Eq. (1). The second one is the mean shift algorithm that implements the approach from Ref. [13], while the quantum clustering uses the mode finding approach from Section 5 to the quantum potential $V(\mathbf{Z})$, whose derivation is explained in Section 3. For all the methods we use the same scale $\hat{\sigma}$, estimated for the corresponding data set, as shown in the plots from Fig. 11. The resulting functional surfaces, representing either $\psi(\mathbf{Z})$ or $V(\mathbf{Z})$ for the two vector fields are shown in Fig. 12. The modes are detected from these functions according to the methodology described in Section 5. The classes corresponding to the detected modes are ordered according to their significance, as given by the factor

L_k , calculated according to Eq. (24). Each mode share of the total energy, L_k , is provided in Fig. 13 for the two data sets, when using the scale-space approach $\psi(\mathbf{Z})$ and the quantum potential $V(\mathbf{Z})$, respectively. It can be observed from Fig. 13 that the quantum potential produces more classes than the scale-space approach using $\psi(\mathbf{Z})$. This is due to the fact that quantum clustering employs the second derivative and is more sensitive to the pdf variation. The knots from the areas corresponding to the quantum potential modes, marked by “*”, are shown for the Wales data in Fig. 10(a) and for the Titan data in Fig. 10(b), respectively.

The SAR image segmentation results based on the orientation of surface normals are displayed in Fig. 14 for the SAR image from Wales and in Fig. 15 for the SAR image from Titan. The 3D terrain surfaces, reconstructed according to the algorithm described in Refs. [30,31], are shown for Wales image in Fig. 14(d), and for Titan data in Fig. 15(d). A total of 5 clusters were found by the scale-space algorithm, 6 by the mean shift and 9 by the quantum clustering for the terrain image from Wales. For the SAR image from Titan, the scale-space algorithm found 8 clusters, mean shift 5, and quantum clustering 10. If we scrutinize the “Wales” data closely, the segmentation around the lake is discernible in the quantum clustering, while with the other two methods is diffused into the area that lies beneath the lake. As it can

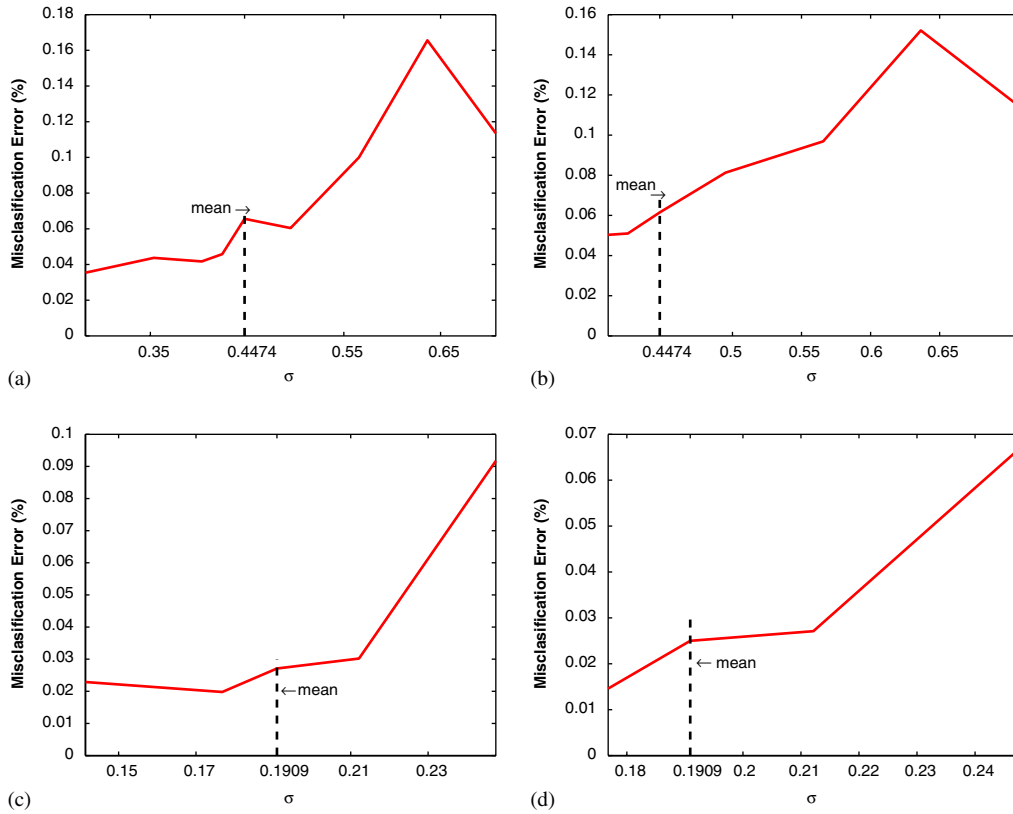


Fig. 8. Classification errors when representing the data assuming various numbers of modes: (a) $\psi(\mathbf{X})$ for 4-QAM; (b) $V(\mathbf{X})$ for 4-QAM; (c) $\psi(\mathbf{X})$ for 8-PSK; (d) $V(\mathbf{X})$ for 8-PSK.

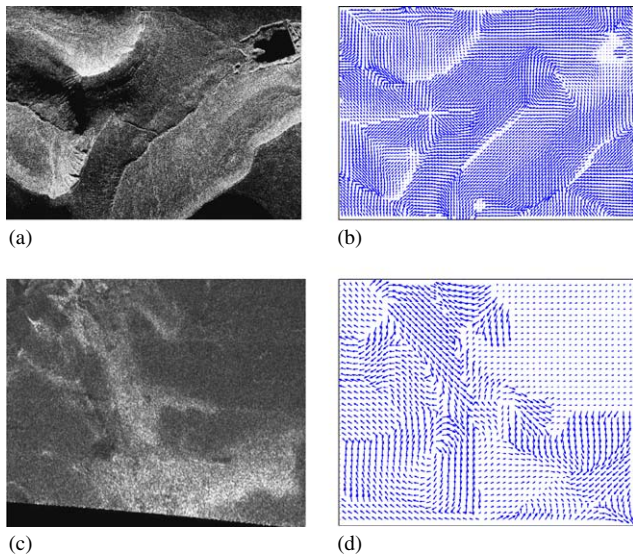


Fig. 9. SAR images of terrain and their corresponding vector fields of surface normals: (a) Wales; (b) smoothed surface normals; (c) Titan; (d) smoothed surface normals.

be observed from these figures, all the regions segmented according to the unsupervised classification of surface normal vectors correspond to compact areas in the original SAR

images. All three nonparametric methods capture the dominant terrain features such as the lake in the top right corner, the slope descending from the semicircular ridge from the left side, as well as the deep valley crossing obliquely the image that ends near the lake from the “Wales” image. A series of terrain features can be clearly identified from the segmentation of the surface normal field extracted from the SAR image from Titan. These results substantiate the capability of the proposed methodology for the classification of vector fields of surface normals according to their orientation. From the surface plots from Fig. 12 we can observe that the scale-space approach given by $\psi(\mathbf{X})$ provides rather shallow local maxima, compared to the local minima produced by the quantum potential $V(\mathbf{X})$. On the other hand, the quantum clustering algorithm consistently produced more clusters for the same data set. This shows its sensitivity to small variations in surface normal orientation.

7. Conclusions

This paper proposes a new methodology for kernel-based nonparametric classification. Two different approaches are considered for modelling the probability density function from a given data set. The first approach consists of the classical kernel density estimation approach, considering

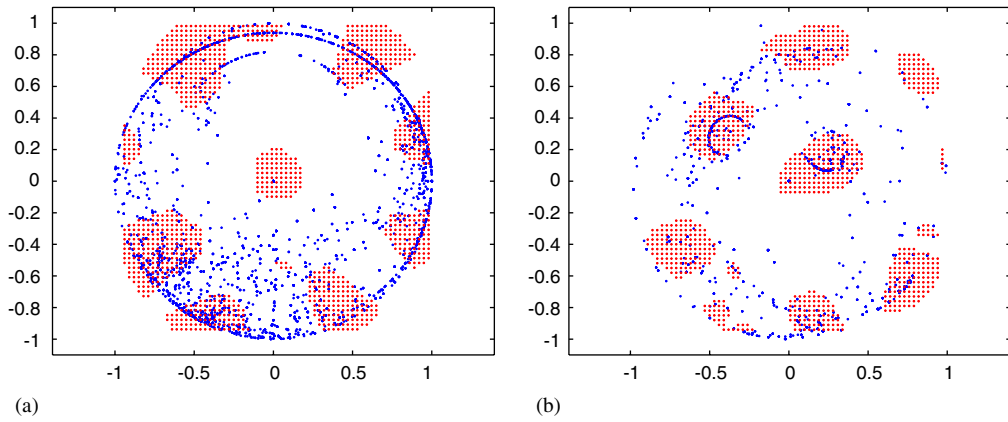


Fig. 10. Data distributions representing surface normal orientation in the (x, y) plane: (a) Wales; (b) Titan.

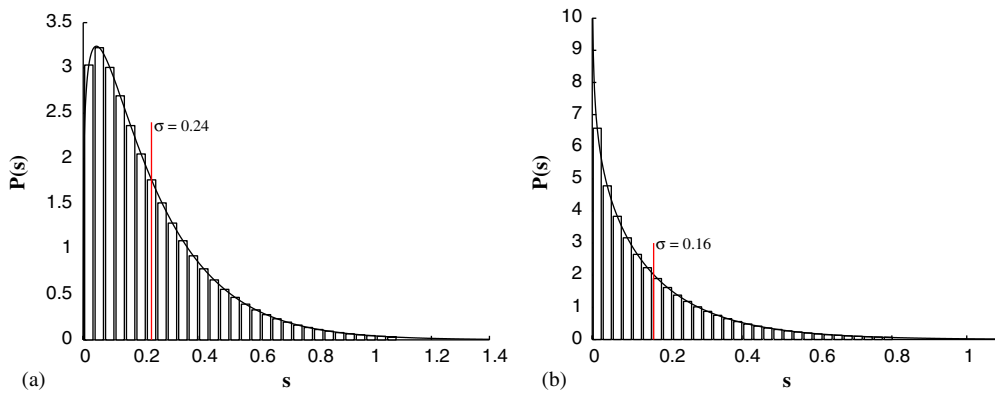


Fig. 11. Gamma fitting to distributions of variance of K -nearest neighbours: (a) Wales; (b) Titan.

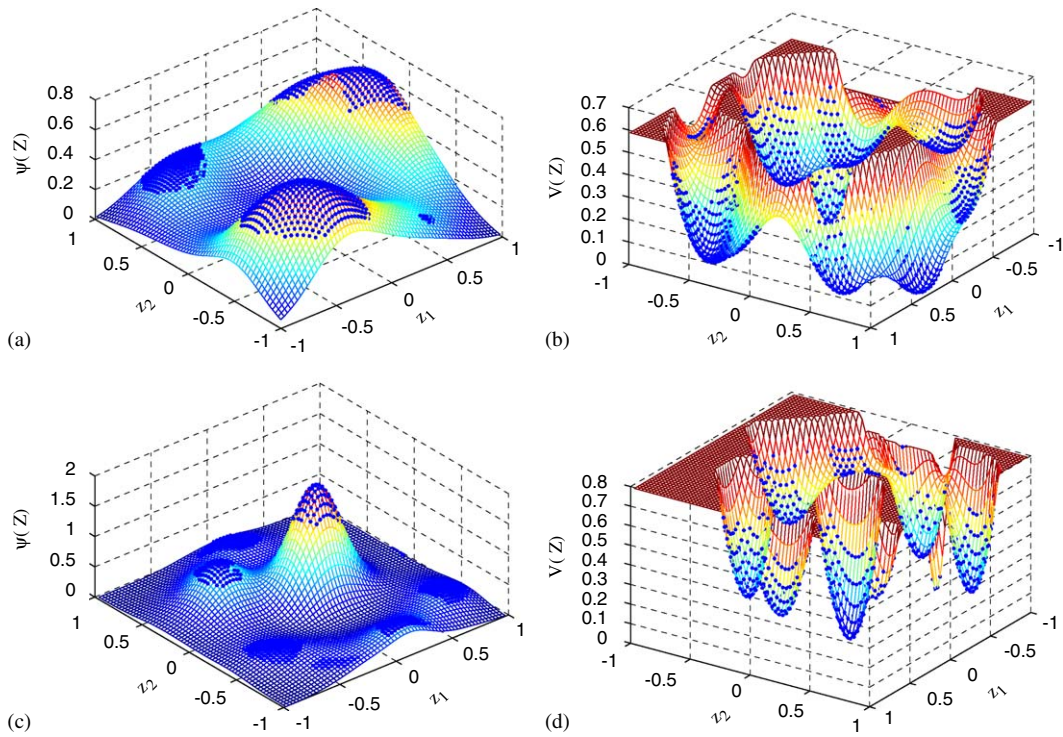


Fig. 12. Functional representations for the vector fields of surface normals. The representation for $V(\mathbf{Z})$ in (b) and (d) is visualized from below in order to evidenciate its modality: (a) $\psi(\mathbf{Z})$ for Wales; (b) $V(\mathbf{Z})$ for Wales; (c) $\psi(\mathbf{Z})$ for Titan; (d) $V(\mathbf{Z})$ for Titan.

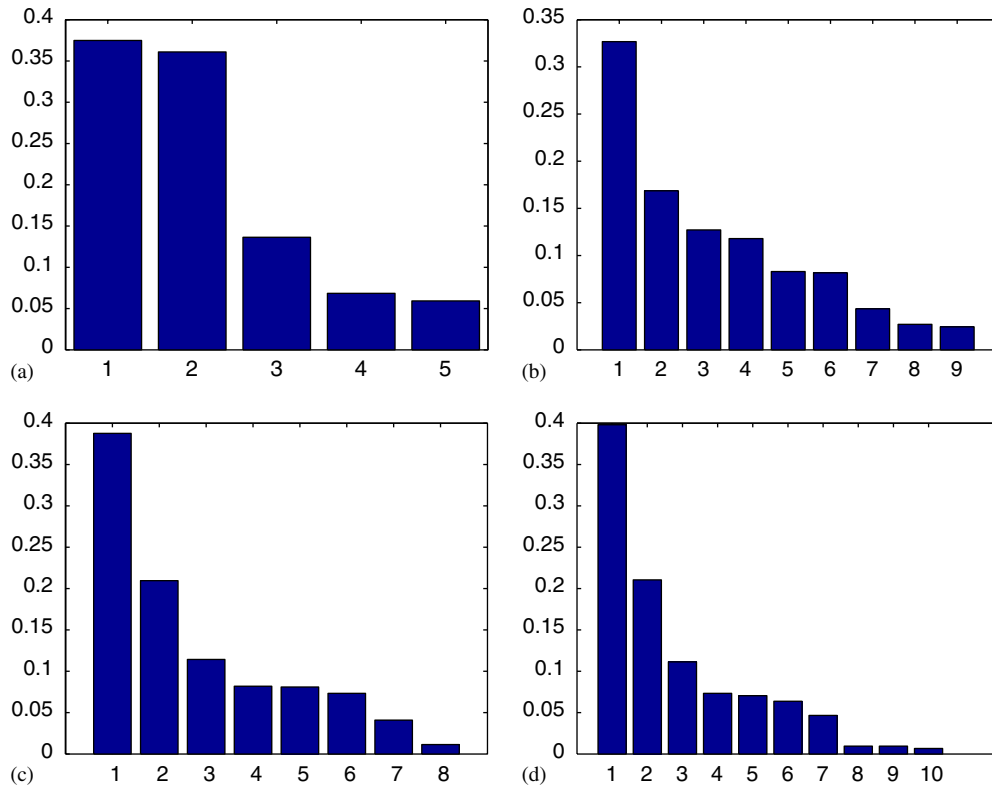


Fig. 13. Ranking of segmented areas according to their share of the potential energy: (a) L_k when using $\psi(\mathbf{X})$ for Wales data; (b) L_k when using $V(\mathbf{X})$ for Wales data; (c) L_k when using $\psi(\mathbf{X})$ for Titan data; (d) L_k when using $V(\mathbf{X})$ for Titan data.

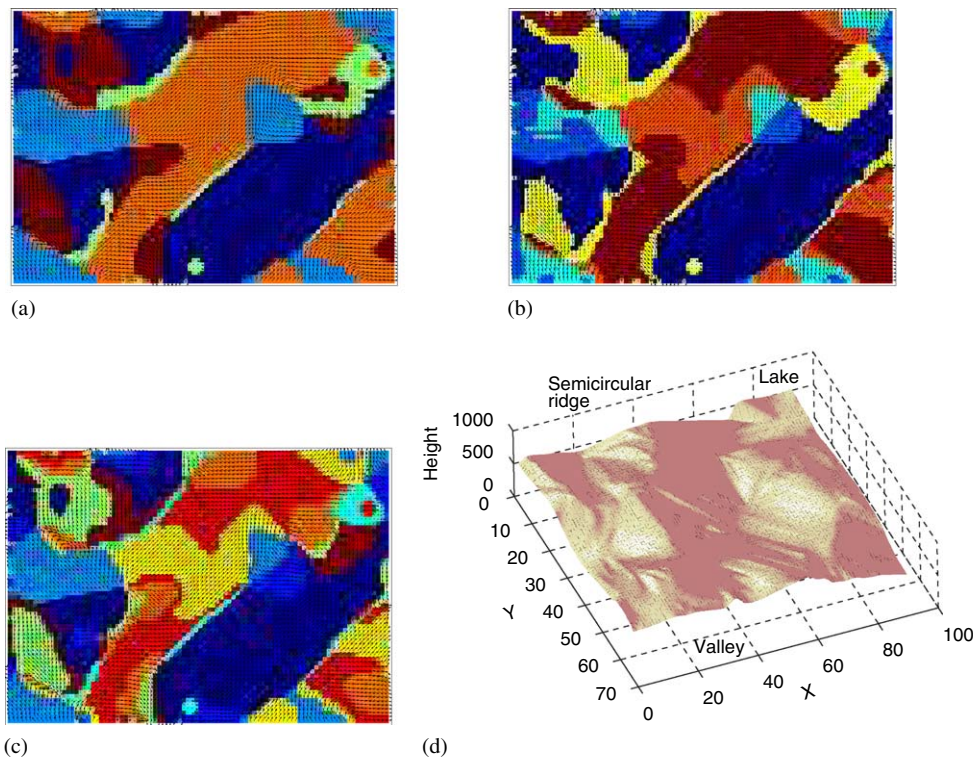


Fig. 14. Topography segmentation of a SAR image of terrain from Wales: (a) scale space; (b) mean shift; (c) quantum clustering; (d) reconstructed 3D map.

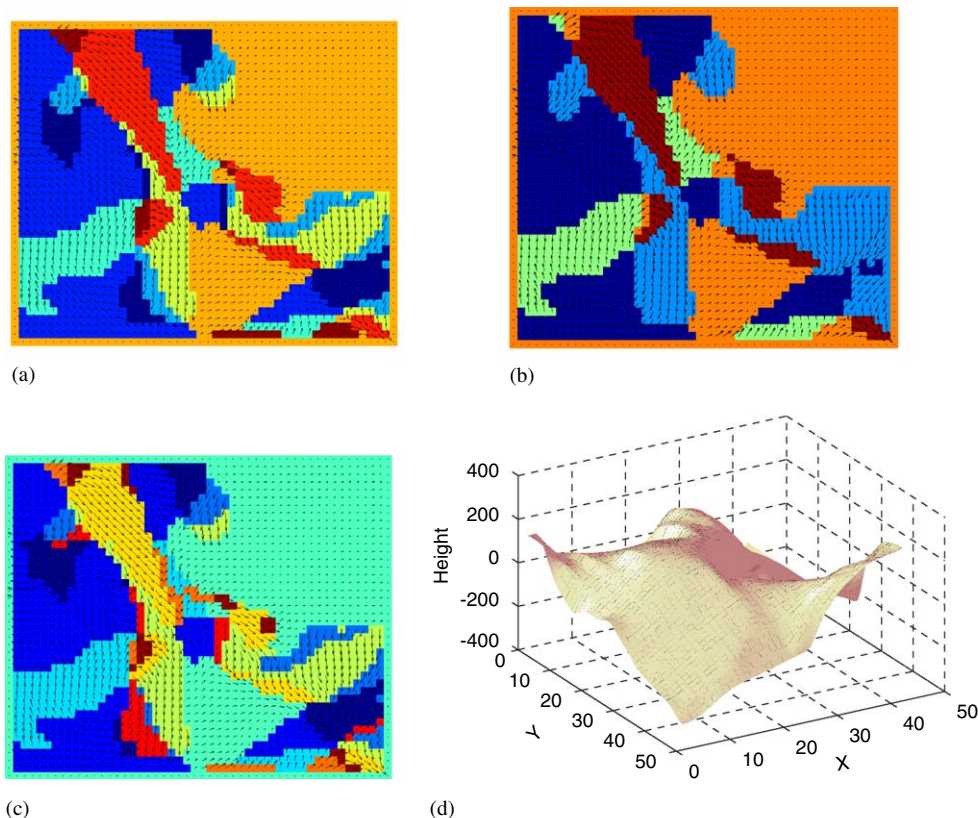


Fig. 15. Topography segmentation of a SAR image from Titan: (a) scale space; (b) mean shift; (c) quantum clustering; (d) estimated 3D reconstruction.

Gaussian kernels, one assigned for each data sample. The second approach considers an algorithm derived from quantum mechanics principles. The algorithm, called quantum clustering, employs the Schrödinger partial differential equation for calculating the quantum potential for a given data set. In quantum mechanics, the Schrödinger equation is used to calculate the probability of locating a particle given its potential energy. In pattern recognition, when using this equation, we consider the reverse problem in which we know the location of a set of data samples and we want to represent their potential function which is assimilated with their characteristic pdf. In our study, the Schrödinger equation acts as a cost function that classifies data according to their similarity. A statistical method that calculates the variance of K -nearest neighbours is used to estimate the kernel parameter, when the kernel function is assumed homoscedastic. The resulting potential function is interpreted using the local Hessian calculated on a lattice. Knots of the lattice are classified as local minima, maxima and as saddle points according to the signs of the local Hessian eigenvalues. By using a region growing algorithm we split the data space in classes without using any parameterized distance measure. The proposed algorithm is applied for separating data from two spirals circling each other, in blind detection of modulated signals and in segmenting vector fields of surface normals extracted from Synthetic Aperture Radar images of terrain. The segmentation of radar images of terrain can be

used in a graph-based representation of topographical information.

References

- [1] R.O. Duda, P.E. Hart, D.G. Stork, *Pattern Classification*, Wiley, New York, 2000.
- [2] S.J. Roberts, Parametric and non-parametric unsupervised cluster analysis, *Pattern Recognition* 30 (2) (1997) 261–272.
- [3] N. Nasios, A.G. Bors, Variational learning for Gaussian mixture models, *IEEE Trans. Syst. Man Cybern.—Part B Cybern.* 36 (4) (2006) 849–862.
- [4] E. Parzen, On estimation of a probability density function and mode, *Ann. Math. Stat.* 33 (1962) 1065–1076.
- [5] B.W. Silverman, *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, London, 1986.
- [6] D.O. Loftsgaarden, C.P. Quesenberry, A nonparametric estimate of a multivariate density function, *Ann. Math. Stat.* 36 (3) (1965) 1049–1051.
- [7] M.C. Jones, J.S. Marron, S.J. Sheather, A brief survey of bandwidth selection for density estimation, *J. Am. Stat. Assoc.* 91 (433) (1996) 401–407.
- [8] D.F. Specht, Probabilistic neural networks and the polynomial adaline as complementary techniques for classification, *IEEE Trans. Neural Networks* 1 (1) (1990) 111–121.
- [9] V. Katkovnik, I. Shmulevich, Kernel density estimation with adaptive varying window size, *Pattern Recognition Lett.* 23 (2002) 1641–1648.
- [10] N.N. Schraudolph, Gradient-based manipulation of nonparametric entropy estimates, *IEEE Trans. Neural Networks* 15 (4) (2004) 828–837.

- [11] K. Fukunaga, L.D. Hostetler, The estimation of the gradient of a density function, with applications in pattern recognition, *IEEE Trans. Inform. Theory* 21 (1) (1975) 32–40.
- [12] Y. Cheng, Mean shift, mode seeking, and clustering, *IEEE Trans. Pattern Anal. Mach. Intell.* 17 (8) (1995) 790–799.
- [13] D. Comaniciu, P. Meer, Distribution free decomposition of multivariate data, *Pattern Anal. Appl.* 2 (1) (1999) 22–30.
- [14] D. Comaniciu, P. Meer, Mean shift: a robust approach toward feature space analysis, *IEEE Trans. Pattern Anal. Mach. Intell.* 24 (5) (2002) 603–619.
- [15] H. Wang, D. Suter, Robust adaptive-scale parametric model estimation for computer vision, *IEEE Trans. Pattern Anal. Mach. Intell.* 26 (11) (2004) 1459–1474.
- [16] M. Blatt, S. Wiseman, E. Domany, Data clustering using a model granular magnet, *Neural Comput.* 9 (8) (1997) 1805–1842.
- [17] D. Horn, Clustering via Hilbert space, *Physica A* 302 (2001) 70–79.
- [18] D. Horn, A. Gottlieb, The method of quantum clustering, in: *Proceedings of Advances in Neural Information Processing Systems (NIPS)*, vol. 14, 2001, pp. 769–776.
- [19] D. Horn, A. Gottlieb, Algorithm for data clustering in pattern recognition problems based on quantum mechanics, *Phys. Rev. Lett.* 88 (1) (2002) 1–4.
- [20] S. Gasiorowicz, *Quantum Physics*, Wiley, New York, 1996.
- [21] H. Haken, H.C. Wolf, *The Physics of Atoms and Quanta*, sixth ed., Springer, Berlin, 2004.
- [22] N. Nasios, A.G. Bors, Nonparametric clustering using quantum mechanics, *Proceedings of the IEEE International Conference on Image Processing*, Genova, Italy, vol. III, 2005, pp. 820–823.
- [23] B.W. Silverman, Choosing the window width when estimating a density, *Biometrika* 65 (1) (1978) 1–11.
- [24] A.C. Cohen, Estimating parameters of Pearson type III populations from truncated samples, *J. Amer. Stat. Assoc.* 45 (251) (1950) 411–423.
- [25] R.M. Haralick, L.G. Shapiro, *Computer and Robot Vision*, Addison-Wesley, Reading, MA, 1992.
- [26] R. Adams, L. Bischof, Seeded region growing, *IEEE Trans. Pattern Anal. Mach. Intell.* 16 (6) (1994) 641–647.
- [27] J. Serra, *Image Analysis and Mathematical Morphology*, Academic Press, New York, 1982.
- [28] S. Singh, 2D spiral pattern recognition with possibilistic measure, *Pattern Recognition Lett.* 19 (2) (1998) 131–147.
- [29] A. Fred, A.K. Jain, Data clustering using evidence accumulation, in: *Proceedings of the International Conference on Pattern Recognition*, Quebec, Canada, vol. IV, 2002, pp. 276–280.
- [30] A.G. Bors, E.R. Hancock, R.C. Wilson, Terrain analysis using radar shape-from-shading, *IEEE Trans. Pattern Anal. Mach. Intell.* 25 (8) (2003) 974–992.
- [31] A.G. Bors, E.R. Hancock, R.C. Wilson, A Bayesian framework for radar shape-from-shading, *Proceedings of the IEEE Computer Vision and Pattern Recognition (CVPR)*, Hilton Head Island, SC, USA, vol. I, 2000, pp. 262–268.
- [32] J. Astola, P. Haavisto, Y. Neuvo, Vector median filters, *Proc. IEEE* 78 (4) (1990) 678–689.

About the Author—NIKOLAOS NASIOS was born in Thessaloniki in 1976. He graduated from the Department of Electrical and Computer Engineering, National Technical University of Athens (NTUA) in 2000 and received his Ph.D. degree in Computer Science from the University of York, UK, in 2006. His research interests lie in the area of image processing, pattern recognition and specifically focus on statistical methods for data clustering. Mr. Nasios is a Member of the Technical Chamber of Greece.

About the Author—ADRIAN G. BORS received his M.S. degree in Electronics Engineering from the Polytechnic University of Bucharest, Romania, in 1992, and his Ph.D. degree in Informatics from the University of Thessaloniki, Greece in 1999. From 1992 to 1993, he was a Research Scientist at the Signal Processing Laboratory, Tampere University of Technology, Finland. From 1993 to 1999 he was a Research Associate, first with the Department of Electrical and Computer Engineering and later on with the Department of Informatics at the University of Thessaloniki, Greece. In March 1999 he joined the Department of Computer Science, University of York, UK, where he is currently a lecturer. Dr. Bors has been an Associate Editor of the *IEEE Transactions on Neural Networks* since 2001, and a Member of the technical committees for various international conferences. He has authored and coauthored 14 journal papers, and more than 50 papers in edited books and international conference proceedings. Dr. Bors research interests include computational intelligence, computer vision, image processing, pattern recognition, digital watermarking, and nonlinear digital signal processing. He is a Senior Member of the IEEE since 2004.