A NEW APPROACH TO DETECTION OF CHANGES IN MULTIDIMENSIONAL PATTERNS

Tomasz Gałkowski[1,*], Adam Krzyżak[2,3], Zbigniew Filutowicz [4,5]

[1]*Institute of Computational Intelligence*
Czestochowa University of Technology
al. Armii Krajowej 36, PL-42-200 Częstochowa, Poland

[2]*Department of Computer Science and Software Engineering*
Concordia University, Montreal, Quebec, Canada H3G 1M8
and Department of Electrical Engineering

[3]*Westpomeranian University of Technology, 70-310 Szczecin, Poland*

[4]*Information Technology Institute*
University of Social Sciences, 90-113 Łódź

[5]*Clark University Worcester, MA 01610, USA*

[*]*E-mail: tomasz.galkowski@pcz.pl*

Abstract

Nowadays, unprecedented amounts of heterogeneous data collections are stored, processed and transmitted via the Internet. In data analysis one of the most important problems is to verify whether data observed or/and collected in time are genuine and stationary, i.e. the information sources did not change their characteristics. There is a variety of data types: texts, images, audio or video files or streams, metadata descriptions, thereby ordinary numbers. All of them changes in many ways. If the change happens the next question is what is the essence of this change and when and where the change has occurred. The main focus of this paper is detection of change and classification of its type. Many algorithms have been proposed to detect abnormalities and deviations in the data. In this paper we propose a new approach for abrupt changes detection based on the Parzen kernel estimation of the partial derivatives of the multivariate regression functions in presence of probabilistic noise. The proposed change detection algorithm is applied to one- and two-dimensional patterns to detect the abrupt changes.

Keywords: edge detection, regression, nonparametric estimation

1 Introduction

Today the huge amounts of varied data are processed, transmitted and stored on the Internet. There is a variety of data types: texts, images, audio or video files or streams, metadata descriptions or just ordinary numbers. The important issue is confidence that these collected data are genuine, or reliable i.e. information sources did not change their characteristics in time. Let us note that we are not

interested in the ordinary distortion of data (errors) by impulse noise, for instance. These types of problems are dealt by error correcting codes and/or digital filtering in telecommunications.

We are rather interested in determining the essence of the changes if they occurred, and where or when they have arisen. The fundamental question is how to detect such changes automatically, i.e. by the appropriate use of a computer program.

There are several methods and algorithms for detecting abnormalities or deviations in the data. The popular approaches are based on the probabilistic and statistical tests or model building. Such strategies are motivated by the lack of precise and complete mathematical description of the process generating the observed data, even if we apply the known physical, biological or medical principles and mathematical equations.

The following general types of changes may be of interest to us (see [2]):
- glitches or anomalies: accidental, fortuitous, often single, less important aberration or simply error in observations caused e.g. by transmission channel noise, temporary disturbances or measurement inaccuracy. Their effect quickly passes away, generally they are irrelevant and can be ignored. It is easy to remove them from the data by appropriate filtering or corrections.
- abrupt or jumping, narrow changes, also called edges: significant deviation from the standard or model observed yet. They are significant to the observer. They may indicate the problem which requires an urgent response. Examples: The abrupt changes in the stock market, abnormalities in physiological parameters of hospital patients, in imaging tomography, changes in geological processes, especially in seismology, cartography, and in several industrial processes. A sudden increase of network traffic flow including the web click-streams, sensor data, phone calls quantities - can indicate possible hacker attack and general lack of information security and the network and system threats. - gradual changes, trends or drifts: subtle shifts that cannot be easily detected manually. They proceed slowly, are barely visible in the long period and more difficult to detect and classify as important. These changes, however, can be of great qualitative importance and may indicate a profound change in the structure of the model. Examples: the changes in

the global temperature of the earth's surface, the amount of underground water resources in certain areas, the gradual general degradation of air quality, changes in the earth magnetic field - all of them, in the long perspective, threaten the human civilization and even life on Earth.

2   A brief overview of the methodologies used to date

There are several methods and algorithms useful to detect abnormalities or deviations in the data.

The survey of the edge detection techniques in *2d*-image processing one can be found in, e.g., [2, 48]. The authors surveyed several approaches for abrupt changes detection via classical gradient-based methods using operations involving first order derivatives such as Sobel, Prewitt, Robert's [31] and Canny [3] edge detectors at which the distribution of intensity values in the neighborhood of given pixels determines the probable edges. The techniques involving second order derivatives such as the Laplacian and Gaussian filtering used for detecting of zero-crossings also allow edge detection in images [29].

In the current studies, the common approach methodology is to conduct phenomena analysis by modelling the problems by multidimensional probability density functions in continuous *d*-dimensional spaces, and distributions in the case of discrete series of random numbers representing statistical multidimensional processes. The more general, natural approach is to model the data via distributions or densities [6]. Using mathematical statistics and the representative templates like means and simple models (e.g. linear regression) the significant features could be compared using different sample sets. This comparison may result in the detection of a change in some parameters, or more general criteria like mean square error can be also used to detect change. When a parameter of the process is estimated the so-called parametric approach is applied.

The nonparametric methodology is used when no assumptions on the functional form on the data have been made. Many statistical tests have been applied in this problem like the Kolmogorov-Smirnov test or Wilcoxon test, for instance (see

[4]). The main idea is to compute a scalar function of the data (so-called test statistics) and compare the values to determine whether a significant change (defined before) has occurred. One of the most general measures of the distance between two distributions is the relative entropy known as the Kullback-Leibler distance [25].

The methods cited above are effective when the volume of the data is not very large, and usable offline. So, for data streams, they are not applicable directly.

The interesting results on various regression models for stream data mining are discussed in [10, 11, 12, 24, 30, 41, 42, 43, 44].

In general, the objects and/or processes can be described mathematically as a function $R(.)$ of the d-dimensional vector variable $\mathbf{X}$. Then the methods based on regression function analysis can be applied. An abrupt change of the function $R(.)$ value at point p may be recognized as a jump discontinuity of the function. In the one dimensional case ($d = 1$) it may be observed as a steep chart. The main problem is to determine the point p at which this occurs. In the case $d > 1$ the place of change (the edge) takes form of a curve in d-dimensional space (across which R is discontinuous). Its calculation is more difficult and requires much more computational effort. One way of detecting change is to compare likelihood between the subsequent examples using adjacent sliding time-windows, for previous elements in the stream and the further ones. The point p could be estimated when we observe a decreasing likelihood.

An interesting solution from the literature applies the Kulback-Leibler divergence e.g. in [14]. The data in consecutive time-windows are clustered using k-means into K clusters. The discrete distribution is calculated where each cluster has a probability proportional to the number of examples it holds. If two distributions are identical the Kullback-Leibler divergence is close to 0, when they are substantially different the Kullback-Leibler divergence is close to 1.

The combined semi-parametric log-likelihood detector, a compromise method between Hoteling (parametric detector) and non-parametric Kullback-Leibler divergence was recently studied in [14] using among others the Mahalanobis distance, and Gaussian mixture of distributions.

Another interesting approach is based on radial basis functions (RBF). The method described in [37] uses the scalable radial kernels in the form $K(\mathbf{x}, \mathbf{y}) := \Phi(\mathbf{x} - \mathbf{y})$ where Φ is a radial function, defined on R^d. It can be rewritten in the form $\Phi(r)$ where $r = \|.\|$ denotes the distance norm and $\Phi : [0, \infty) \rightarrow \Re$ - a function of a single non-negative real variable. The authors have chosen the Wendland kernels of polynomials with even order of smoothness. Kernels on R^d can be scaled by the positive factor delta in the following way: $K(\mathbf{x}, \mathbf{y}, \delta) := K\left(\frac{\mathbf{x}}{\delta}, \frac{\mathbf{y}}{\delta}\right), \forall \mathbf{x}, \mathbf{y} \in R^d$.

The parameter δ is called the shape or scale parameter and can be tuned by the experimenter (depending to the application). It controls the accuracy of the interpolation and the stability. The main idea is to interpolate the data with the radial kernel functions, next calculating the set of the coefficients of this interpolation in some cardinal function. The main idea is based on the so-called Gibbs phenomenon: when the approximated function has the discontinuity at the point p in the Fourier series near this point, the high frequency components arise, so the corresponding Fourier coefficients take larger absolute values in this region. A suitable thresholding strategy could be used to detect the point p.

In this paper, we focus our attention on the challenge of abrupt change detection, also called edge detection problem, by presenting the new original approach. The main result is the method of edge detection derived from the nonparametric approach based on Parzen kernel algorithms for estimation of unknown regression functions and their derivatives from the set of noisy measurements. The algorithms are developed for two-dimensional functions or patterns on 2d plane. Restricting our considerations to *2-dimensional* space allows to better understand the proposed approach, but by no means precludes its generalization to *d-dimensional* space.

This article describes techniques applicable in the wide range of fields such as classification, computer vision, diagnostics etc. (see e.g. [7, 8, 20, 21, 22, 49]). The approach based on regression analysis is developed as an attractive tool also in classification and modelling of objects (e.g. [27, 28]), forecasting of phenomena (e.g. [26, 36]), and entire methodology of machine learning like neural networks, fuzzy sets, genetic algorithms (e.g.

[5, 45, 46]). Nonparametric approach to analysis and modelling of various systems one may found e.g. in [34, 35, 38, 39, 40]. Edge detection technique based on Parzen kernel estimate has also been described by Qiu in [32, 33]. Unlike our algorithm the algorithm presented in [33] is quite complicated and its performance in real applications has not been investigated. The algorithm described in [32] is significantly different from our approach and it uses derivatives computed in a very inefficient way. The algorithm has not been thoroughly tested in experiments. In our approach we compute the derivatives of the kernel itself which is a very simple and efficient process and our algorithms performs in satisfactory manner in numerical experiments. Furthermore, our algorithm can scale up and it does not require the samples to be uniformly spaced.

3　New concept of the algorithm for abrupt change detection - one-dimensional case

The RBF methods described in Section 2 are of the kernel-type methods.

The main goal of this paper is to introduce a new simple method of edge detection derived from the nonparametric approach based on multidimensional Parzen kernel algorithms for estimating unknown functions and their derivatives from the set of noisy measurements. See [23] for theoretical analysis of Parzen and other nonparametric regression estimation techniques for so-called random design case.

We consider the model of the object in the form

$$y_i = R(\mathbf{x}_i) + \varepsilon_i, \quad i = 1, \dots, n, \quad (1)$$

where $\mathbf{x_i}$ is assumed to be the d-dimensional vectors of deterministic input, $\mathbf{x}_i \in R^d$, y_i is the scalar random output, and ε_i is a measurement noise with zero mean and bounded variance. $R(.)$ is assumed to be completely unknown function. This is so-called fixed-design regression problem, see e.g. [13].

We start with estimator $\hat{R}_n(\mathbf{x})$ of function $R(.)$ at point $\mathbf{x}$ based on the set of measurements y_i, $i = 1, \dots, n$. We use the Parzen kernel based algorithm of the integral type

$$\hat{R}(\mathbf{x}) = h_n^{-d} \sum_{i=1}^{n} y_i \int_{D_i} \mathbf{K}\left(\frac{\|\mathbf{x}-\mathbf{u}\|}{h_n}\right) d\mathbf{u}, \quad (2)$$

where $\|\mathbf{x}-\mathbf{u}\|$ denotes a norm or the distance function defined for points $\mathbf{x}$ and $\mathbf{u}$ in d-dimensional space and D_i's are defined below.
Factor h_n depending on the number of observations n is called the smoothing factor.

Let us mention that in nonparametric approach we impose no constraints on either the shape of unknown function (like e.g. in the spline methods or linear regression) or on any mathematical formula with a certain set of parameters to be found (like in so-called parametric approach).

The domain area D (the space where function R is defined) is partitioned into n disjunctive nonempty sub-spaces D_i and the measurements $\mathbf{x}_i$ are chosen from D_i, i.e.: $\mathbf{x}_i \in D_i$.
For instance, in one-dimensional case let the $D = [0,1]$, then $\cup D_i = [0,1]$, $D_i \cap D_j = \emptyset$ for $i \neq j$, the points x_i are chosen from D_i, i.e.: $x_i \in D_i$.

The set of input values $\mathbf{x}_i$ (independent variable in the model (1) are chosen in the process of collecting data e.g., equidistant samples of ECG signal in time domain, or stock exchange information, or internet activity on specified TCP/IP port of the web or ftp server logs recorded in time. These data points should provide a balanced representation of function R in the domain D.
The standard assumption in theorems on convergence of (3) is that the maximum diameter of set D_i tends to zero if n tends to infinity (see e.g. [18, 15, 16]). We may assume that in the set of pairs $(\mathbf{x_i}, \mathbf{y_i})$ information (in some way *inscribed*) on essential properties of function R, like its smoothness is present.

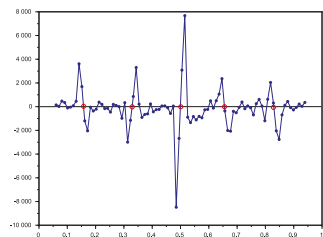The kernel function $\mathbf{K}$ in one-dimensional case $K(.)$ satisfies the following conditions

$$
\begin{aligned}
K(t) &= 0 \qquad t \notin (-\tau, \tau), \tau > 0 \\
\int_{-\tau}^{\tau} K(t)dt &= 1 \\
\sup_t |K(t)| &< \infty.
\end{aligned}
\qquad (3)
$$
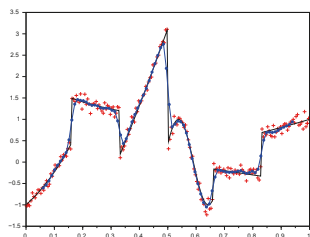
Estimated function (6) with added noise
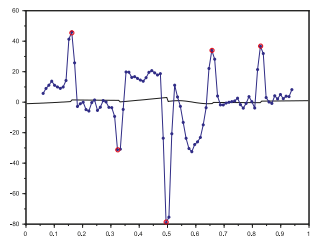
First derivative of function (6)

Second derivative of function (6)

Figure 1. Simulation example - function (6)

We will use the following trigonometric cosine kernel satisfying (3)

$$K(t) = \begin{cases} \frac{\pi}{4}\cos\left(\frac{\pi}{2}t\right) & for \quad t \in (-1,1) \\ 0 & otherwise \end{cases}. \quad (4)$$

The algorithm for estimating the derivatives of order k is based on differentiation of the kernel function. Thus the kernel $K(.)$ must be differentiable function of order k. The trigonometric cosine kernel (4) fulfils this condition. The estimate of $k-th$ derivative of the regression function in point x can be defined as follows

$$\hat{R}^{(k)}(x) = h_n^{-1}\sum_{i=1}^{n} y_i \int_{D_i} K^{(k)}\left(\frac{x-u}{h_n}\right) du \quad (5)$$

The nonparameric approach in application to estimation of unknown functions and their derivatives was previously proposed and studied in univariate case in e.g. [19, 17].

The main idea of the paper is to deduce the dynamics of changes of any function by analysing the course of the first derivative estimated from sample. The more rapidly the change occurs - the higher the first derivative (or speed). The steeper the slope - the larger the tangent referring to horizon surface at a given point. These facts motivate us to propose as a detector of abrupt changes the nonparametric estimator of the derivatives described previously. The smoothing property of these algorithms makes it applicable when the observations are taken in the presence of random noise. The integral version of nonparametric kernel estimation algorithms (2), using the Parzen kernel (3), enables not only estimation of the value of the desired regression function, but also estimation of the value of its first derivative and higher order derivatives, too.

The appropriate strategy using the estimates of the first and the second derivatives of R for edge detection is proposed. Using only the first derivative we need the appropriate thresholding strategy to detect jumps in function R, however, applying the second derivative and finding its zero values lets us to determine edges directly.

Next we present simulation results. The function chosen for testing is a slightly modified function proposed first by Romani et al. in [37] and is defined as follows

$$f(x) = \begin{cases} (1+x)^6 - 2 & 0 \le x < 0.16 \\ (1-x)^4 + 1 & 0.16 \le x < 0.33 \\ (2+x)^3 + 12.5 & 0.33 \le x < 0.5 \\ \sin(30x-2) & 0.5 \le x < 0.66 \\ -x+0.5 & 0.66 \le x < 0.83 \\ x^2 & 0.83 \le x \le 1 \end{cases}. \quad (6)$$

Figure 1. presents the simulation results for noised inputs generated from equation (6). Diagrams show the original function (6) and its 1-st and 2-nd derivative, respectively. The black solid lines - functions and derivatives, their nonparametric estimates - blue points and lines. Red pluses represent the noised measurements. The maxima of the first derivative (marked as red points) assign the points

of possible jumps or abrupt changes. In the third row, we have the second derivative. By observing the zeros of this function in relation to the corresponding maxima of the first derivative we may deduce at which points the jumps have occurred.

The smoothing properties of the nonparametric algorithm of the Parzen kernel type depend on the parameter h_n. The choice of its value plays an important role in the interpretation of results. The bigger the h_n the bigger the level of smoothness, but then detection at which point the jump has occurred is more difficult. On the other hand, a too small value of h_n causes higher oscillations of the estimates of the derivatives and consequently, the bigger number of sharp peaks of the first derivative. Optimal choice of smoothing sequence or bandwidth is rather difficult and it is often data dependent, see [50, 9, 1].

4 Main results - the multidimensional algorithm of abrupt change detection
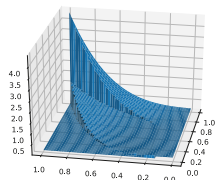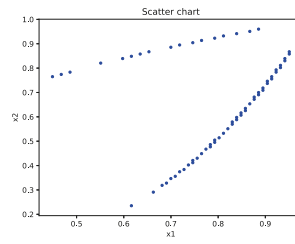

Original function without noise (15)
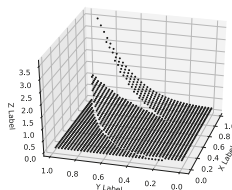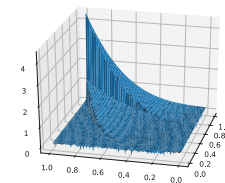

Function (15) probe set
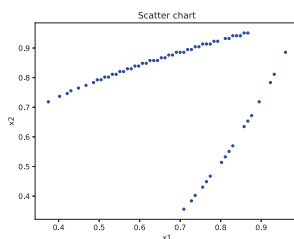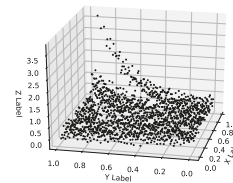

Edge curves - the scatter diagram No. 1

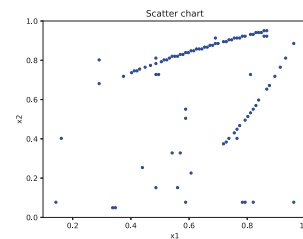Edge curves - the scatter diagram No. 2

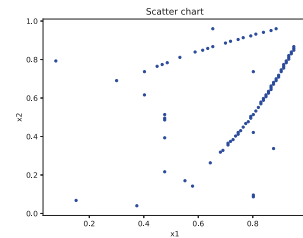**Figure 2**. Simulation Example 1 - function (15)

Original function with noise (15)

Function (15) noised probe set

Edge curves - the scatter diagram No. 1

Edge curves - the scatter diagram No. 2

Figure 3. Simulation Example 2 - function (15) with noise

In this Section, we propose the multidimen-

sional extension of the nonparametric method of edge detection. In the multidimensional case there are two commonly used approaches:
- the radial kernel (using e.g. the norm in Euclidean sense) given by

$$\mathbf{K}\left(\mathbf{u}^T\mathbf{u}\right) = c \cdot \sqrt{\mathbf{u}^T\mathbf{u}}, \qquad (7)$$

- the product kernel given by

$$\mathbf{K}\left(\mathbf{x},\mathbf{u},h_n\right) = \prod_{p=1}^{d} K\left(\frac{|x_p - u_p|}{h_n}\right) = \mathbf{K}\left(\frac{\|\mathbf{x} - \mathbf{u}\|}{h_n}\right). \qquad (8)$$

The computationally more efficient is the radial kernel, but the simplicity and ease of application of product kernels make them more preferred in practice, particularly when differentiation is needed. We use the product kernel (8) in this paper.

In the multidimensional case $(d > 1)$ the estimate of partial derivative of order k with respect to the coordinate variable x_j is given by

$$\hat{R}_{x_j}^{(k)}(\mathbf{x}) = h_n^{-d}\sum_{i=1}^{n} y_i \int_{D_i} \frac{\partial^k}{\partial x_j^k}\mathbf{K}\left(\frac{\|\mathbf{x} - \mathbf{u}\|}{h_n}\right)d\mathbf{u}. \qquad (9)$$

It is clear that the estimation of particular derivative is obtained by the differentiation of the kernel function depending on the relative coordinate. Let us analyze the two-dimensional case.
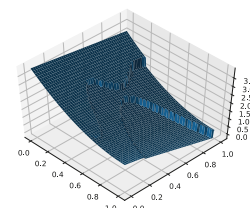
The model of the object is now in the form

$$y_i = R([x_1,x_2]_i) + \varepsilon_i, \quad i = 1,...,n, \qquad (10)$$
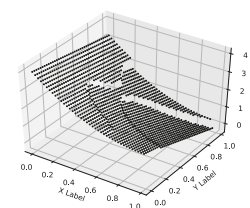
where the $2d$-vector of independent variable: $\mathbf{x}_i = [x_1,x_2]_i$.
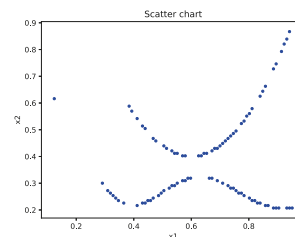The $2d$ Parzen kernel based estimator is defined by

$$\hat{R}([x_1,x_2]) = h_n^{-2}\sum_{i=1}^{n} y_i \cdot$$
$$\cdot \int_{D_i} K\left(\frac{x_1 - u_1}{h_n}\right) \cdot K\left(\frac{x_2 - u_2}{h_n}\right)du_1 du_2. \qquad (11)$$
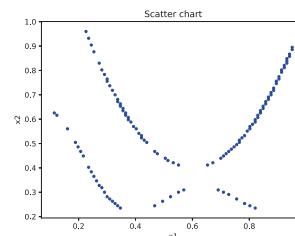

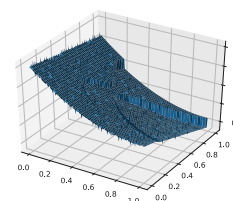Original function (16)

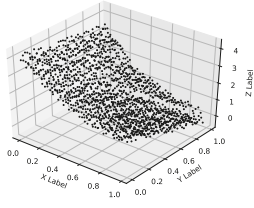Function (16) probe set

Edge curves - the scatter diagram No. 1

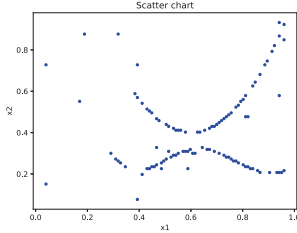Edge curves - the scatter diagram No. 2

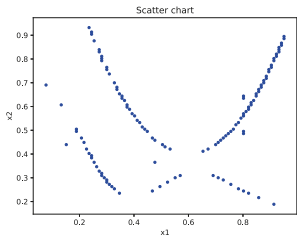**Figure 4**. Simulation Example 3 - function (16)

Original function with noise (16)

Function (16) noised probe set

Edge curves - the scatter diagram No. 1

Edge curves - the scatter diagram No. 2

Figure 5. Simulation Example 4 - function (16) with noise

By using the cosine kernel defined in unidimensional case by (4) we obtain the estimator based on product kernel in the form

$$\hat{R}([x_1,x_2]) = \frac{\pi^2}{16} \cdot h_n^{-2} \sum_{i=1}^{n} y_i \cdot \\ \cdot \int_{D_i} \cos\left(\frac{\pi(x_1-u_1)}{2h_n}\right) \cdot \cos\left(\frac{\pi(x_2-u_2)}{2h_n}\right) du_1 du_2 \quad (12)$$

Now we can derive the estimators of the partial derivatives on coordinates x_1 and x_2, respectively - as follows

$$\frac{\partial}{\partial x_1}\hat{R}([x_1,x_2]) = \frac{\pi^3}{32} \cdot h_n^{-3} \sum_{i=1}^{n} y_i \int_{D_i} \sin\left(\frac{\pi(u_1-x_1)}{2h_n}\right) \cdot \\ \cdot \cos\left(\frac{\pi(x_2-u_2)}{2h_n}\right) du_1 du_2 \quad (13)$$

$$\frac{\partial}{\partial x_2}\hat{R}([x_1,x_2]) = \frac{\pi^3}{32} \cdot h_n^{-3} \sum_{i=1}^{n} y_i \cdot \\ \cdot \int_{D_i} \cos\left(\frac{\pi(x_1-u_1)}{2h_n}\right) \cdot \sin\left(\frac{\pi(u_2-x_2)}{2h_n}\right) du_1 du_2. \quad (14)$$

The integrals in (13) and (14) are easy to calculate analytically.

The last requirement is to prepare the subsets D_i keeping in mind that the points x_i should be chosen from D_i, i.e.: $x_i \in D_i$. The natural solution is an equally spaced grid easy to construct but it is not detailed here.

5   Simulation results

A series of simulation tests were carried out to assess the effectiveness of the proposed algorithms. Two groups of tests were performed: at first, tests based on generated sets of measurement pairs $\mathbf{x}, \mathbf{y}$ without measurement noise, and second, the sets of measurements with additive noise.

Two functions are considered. The first is the two-dimensional original function from Romani et al. [37] given by

$$R_1(x) = \begin{cases} (x_1+x_2-0.5)^3 \cdot x_1 & for & x_1 \leq x_2^3 \\ (x_1+x_2)^{1/3} \cdot x_1^3 + 0.2 & for & x_1^3 \leq x_2 \\ 0 & else \end{cases} . \quad (15)$$

The second function is defined by equations

$$R_2(x) = \begin{cases} \left(3-1.5x_1+5(x_2-1)^2\right)/2 \\ \qquad\qquad for\ x_2 \leq L1 \\ \left(3-1.5x_1+5(x_2-1)^2\right)/2-0.5x_1 \\ \qquad\qquad for\ L1 \leq x_2 \leq L2 \\ \left(3-1.5x_1+5(x_2-1)^2\right)/2-x_1 \\ \qquad\qquad elsewhere \end{cases} , \quad (16)$$

where

$$L1 = 0.2 + 5 \cdot (0.175 + 0.15\sin(12x_1)) \cdot \\ \cdot (0.1/(x_1+0.5)^2) \quad (17) \\ L2 = 0.4 + 4 \cdot (x_1 - 0.6)^2$$

The simulation results are given in Figures 2-6. Figure 2. shows function (15) without noise, its

probe set (200x200 points) and edge curves (scatter diagrams) obtained by using derivatives with respect to x_1 and x_2, respectively. Figure 3. presents function (15) with additive random noise, its probe set, and edge curves (scatter diagrams) obtained by differentiation on x_1 and x_2, respectively. In the next Figures 4 and 5 analogical simulation results are shown for function defined by (16). Figure 4. shows: the original function (16) without noise, its probe set (200x200 points) and scatter diagrams of estimated edge curves obtained by differentiation with respect to x_1 and x_2. Figure 5. presents the results for function (16) with additive noise. Random noise was generated using uniform distribution from the interval $[-0.5, 0.5)$. The smoothing parameter was experimentally established and its chosen value was $h_n = 0.03$. The application of the algorithm for detection of edges in black-and-white picture was demonstrated in Figure 6. The dotted contours of geometrical symbols are shown in the subsequent diagrams.

Black-and-white sample picture
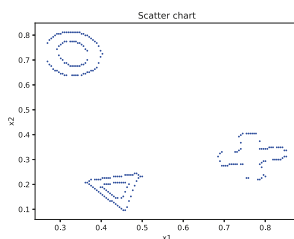
Edges detected by differentiation on x1-coordinate
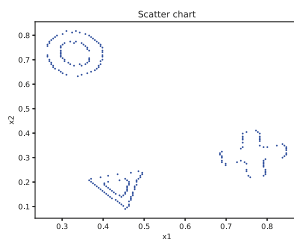
Edges detected by differentiation on x2-coordinate

Figure 6. Simulation Example 5 - Sample black-and-white picture edge detection

Conclusion

This paper considered the important problem of deciding whether the abrupt or sudden change occurred in the data and where or when it happened. The proposed algorithm is derived from the nonparametric kernel regression estimation techniques with fixed-design of unknown functions and their partial derivatives in multidimensional space. The proposed two-dimensional algorithm is presented in detail. The algorithm is tested for jump detection in case of measurements generated without and in the presence of random noise. We also studied the performance of the algorithm in the problem of edge detection in black-and-white images. Simulation results shown in the series of diagrams confirmed utility of the proposed approach in practical cases. From the Figures presented one may observe that the effectiveness of the method improves when the magnitude of the jump is higher. The extension of the edge detection algorithm to multivariate ($d > 2$) case is described and it directly follows from the presented methodology using the product-type of multidimensional kernel.

References

[1] A. Berlinet, G. Biau, L. Rouviere, Optimal L1 bandwidth selection for variable kernel density estimates, Statistics and Probability Letters, Elsevier, Vol. 74, No. 2, 2005, pp. 116-128.

[2] S. Bhardwaj, A. Mittal, A survey on various edge detector techniques, Elseiver, SciVerse ScienceDirect, Procedia Technology 4, 2nd International Conference on Computer, Communication, Control and Information Technology, 2012, pp. 220-226.

[3] J.F. Canny, A computational approach to edge detection, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 8, No. 6, 1986, pp. 679-698.

[4] G.W. Corder, D.I. Foreman, Nonparametric Statistics: A Step-by-Step Approach. Wiley, New York, 2014.

[5] K. Cpałka, L. Rutkowski, Evolutionary learning of flexible neuro-fuzzy systems, Proc. of the 2008 IEEE Int. Conference on Fuzzy Systems (IEEE World Congress on Computational Intelligence, WCCI 2008), Hong Kong June 1-6, CD, 2008, pp. 969-975.

[6] T. Dasu, S. Krishnan, S. Venkatasubramanian, K. Yi, An information-theoretic approach to detecting

changes in multi-dimensional data streams, Proc. Symp. on the Interface of Statistics, Computing Science, and Applications, 2006.

[7] J.J. Davis, Ch.-T. Lin, G. Gillett, R. Kozma, An integrative approach to analyze EEG signals and human brain dynamics in different cognitive states, Journal of Artificial Intelligence and Soft Computing Research, Vol. 7, No. 4, 2017, pp. 287-299.

[8] V.S. Devi, L. Meena, Parallel MCNN (PMCNN) with application to prototype selection on large and streaming data, Journal of Artificial Intelligence and Soft Computing Research, Vol. 7, No. 3, 2017, pp. 155-169.

[9] L. Devroye, G. Lugosi, Combinatorial Methods in Density Estimation. Springer-Verlag, New York, 2001.

[10] P. Duda, M. Jaworski, L. Rutkowski, Convergent time-varying regression models for data streams: tracking concept drift by the recursive Parzen-based generalized regression neural networks, International Journal of Neural Systems, Vol. 28, No. 2, 1750048, 2018.

[11] P. Duda, M. Jaworski, L. Rutkowski, Knowledge discovery in data streams with the orthogonal series-based generalized regression neural networks, Information Sciences, Vol. 460-461, 2018, pp. 497-518.

[12] P. Duda, L. Rutkowski, M. Jaworski, D. Rutkowska, On the Parzen kernel-based probability density function learning procedures over time-varying streaming data with applications to pattern classification, IEEE Transactions on Cybernetics, 2018, pp. 1-14.

[13] R.L. Eubank, Nonparametric Regression and Spline Smoothing. 2nd edition, Marcel Dekker, New York, 1999.

[14] W.J. Faithfull, J.J. Rodríguez, L.I. Kuncheva, Combining univariate approaches for ensemble change detection in multivariate data, Elseiver, Information Fusion, Vol. 45, 2019, pp. 202-214.

[15] T. Gałkowski, L. Rutkowski, Nonparametric recovery of multivariate functions with applications to system identification, Proceedings of the IEEE, Vol. 73, 1985, pp. 942-943.

[16] T. Gałkowski, L. Rutkowski, Nonparametric fitting of multivariable functions, IEEE Transactions on Automatic Control, Vol. AC-31, 1986, pp. 785-787.

[17] T. Gałkowski, On nonparametric fitting of higher order functions derivatives by the kernel method - a simulation study, Proceedings of the 5-th Int.

Symp. on Applied Stochastic Models and data Analysis, Granada, Spain, 1991, pp. 230-242.

[18] T. Gasser, H.-G. Müller, Kernel estimation of regression functions, Lecture Notes in Mathematics, Vol. 757. Springer-Verlag, Heidelberg, 1979, pp. 23-68.

[19] T. Gasser, H.-G. Müller, Estimating regression functions and their derivatives by the kernel method, Scandinavian Journal of Statistics, Vol. 11, No. 3, 1984, pp. 171-185.

[20] R. Grycuk, R. Scherer, M. Gabryel, New image descriptor from edge detector and blob extractor. Journal of Applied Mathematics and Computational Mechanics, Vol. 14, No.4, 2015, pp. 31-39.

[21] R. Grycuk, M. Knop, S. Mandal, Video key frame detection based on SURF algorithm. International Conference on Artificial Intelligence and Soft Computing, ICAISC'2015, Springer, Cham, 2015, pp. 566-576.

[22] R. Grycuk, M. Gabryel, M. Scherer, S. Voloshynovskiy, Image descriptor based on edge detection and crawler algorithm. In International Conference on Artificial Intelligence and Soft Computing, ICAISC'2016, Springer, 2016, pp. 647-659.

[23] L. Györfi, M. Kohler, A. Krzyzak, H. Walk, A Distribution-Free Theory of Nonparametric Regression. Springer, 2002.

[24] M. Jaworski, P. Duda, L. Rutkowski, New splitting criteria for decision trees in stationary data streams, IEEE Transactions on Neural Networks and Learning Systems, Vol. 29, No. 6, 2018, pp. 2516-2529.

[25] S. Kullback, R.A. Leibler, On information and sufficiency, The Annals of Mathematical Statistics. Vol. 22, No. 1, 1951, pp. 79-86.

[26] M.W.Y. Lam, One-match-ahead forecasting in two-team sports with stacked Bayesian regressions, Journal of Artificial Intelligence and Soft Computing Research, Vol. 8, No. 3, 2018, pp. 159-171.

[27] K. Łapa, K. Cpałka, A. Przybył, K. Grzanek, Negative space-based population initialization algorithm (NSPIA), Artificial Intelligence and Soft Computing, ICAISC'2018, Lecture Notes in Computer Science, Vol. 10841, Springer, 2018, pp. 449-461.

[28] K. Łapa, K. Cpałka, A. Przybył, Genetic programming algorithm for designing of control systems, Information Technology and Control, vol. 47, no. 5, 2018, pp. 668-683.

[29] D. Marr, E. Hildreth, Theory of edge detection, Proc. R. Soc. London, B-207, 1980), pp. 187-217.

[30] L. Pietruczuk, L. Rutkowski, M. Jaworski, P. Duda, How to adjust an ensemble size in stream data mining?, Information Sciences, Elsevier Science Inc., Vol. 381, No. C, 2017, pp. 46-54.

[31] W.K. Pratt, Digital Image Processing, 4th Edition, John Wiley Inc., New York, 2007.

[32] P. Qiu, Nonparametric estimation of jump surface, The Indian Journal of Statistics, Series A, Vol. 59, No. 2, 1997, pp. 268-294.

[33] P. Qiu, Jump surface estimation, edge detection, and image restoration, Journal of the American Statistical Association, No. 102, 2007, pp. 745-756.

[34] E. Rafajłowicz, R. Schwabe, Halton and Hammersley sequences in multivariate nonparametric regression, Statistics and Probability Letters, Vol. 76, No. 8, 2006, pp. 803-812.

[35] W. Rafajłowicz, Nonparametric estimation of continuously parametrized families of probability density functions – Computational aspects, Preprint of the Department of Engineering Informatics, Wrocław University of Science and Technology, Wrocław, 2020.

[36] C.R. Rivero, J. Pucheta, S. Laboret, V. Sauchelli, D. Patino, Energy associated tuning method for short-term series forecasting by complete and incomplete datasets, Journal of Artificial Intelligence and Soft Computing Research, Vol. 7, No. 1, 2017, pp. 5-16.

[37] L. Romani, M. Rossini, D. Schenone, Edge detection methods based on RBF interpolation, Journal of Computational and Applied Mathematics, Vol. 349, 2019, pp. 532-547.

[38] L. Rutkowski, Application of multiple Fourier-series to identification of multivariable non-stationary systems, International Journal of Systems Science, Vol. 20, No. 10, 1989, pp. 1993-2002.

[39] L. Rutkowski, E. Rafajłowicz, On optimal global rate of convergence of some nonparametric identification procedures, IEEE Transactions on Automatic Control, Vol. 34, No. 10, 1989, pp. 1089-1091.

[40] L. Rutkowski, Identification of MISO nonlinear regressions in the presence of a wide class of disturbances, IEEE Transactions on Information Theory, Vol. 37, No. 1, 1991, pp. 214-216.

[41] L. Rutkowski, L. Pietruczuk, P. Duda, M. Jaworski, Decision trees for mining data streams based on the McDiarmid's bound, IEEE Transactions on Knowledge and Data Engineering, Vol. 25, No. 6, 2013, pp. 1272-1279.

[42] L. Rutkowski, M. Jaworski, L. Pietruczuk, P. Duda, Decision trees for mining data streams based on the Gaussian approximation, IEEE Transactions on Knowledge and Data Engineering, Vol. 26, No. 1, 2014, pp. 108-119.

[43] L. Rutkowski, M. Jaworski, L. Pietruczuk, P. Duda, The CART decision tree for mining data streams, Information Sciences, Vol. 266, 2014, pp. 1-15.

[44] L. Rutkowski, M. Jaworski, L. Pietruczuk, P. Duda, A new method for data stream mining based on the misclassification error, IEEE Transactions on Neural Networks and Learning Systems, Vol. 26, No. 5, 2015, pp. 1048-1059.

[45] T. Rutkowski, J. Romanowski, P. Woldan, P. Staszewski, R. Nielek, L. Rutkowski, A content-based recommendation system using neuro-fuzzy approach, International Conference on Fuzzy Systems: FUZZ-IEEE, 2018, pp. 1-8.

[46] T. Rutkowski, J. Romanowski, P. Woldan, P. Staszewski, R. Nielek, Towards interpretability of the movie recommender based on a neuro-fuzzy approach, Lectures Notes in Artificial Intelligence, ICAISC'2018, Vol. 10842, Springer, 2018, pp. 752-762.

[47] L. Rutkowski, M. Jaworski, P. Duda, Stream Data Mining: Algorithms and Their Probabilistic Properties, Springer, 2019.

[48] S. Singh, R. Singh, Comparison of various edge detection techniques, in: 2nd International Conference on Computing for Sustainable Global Development, 2015, pp. 393-396.

[49] T. Tezuka, Ch. Claramunt, Kernel analysis for estimating the connectivity of a network with event sequences, Journal of Artificial Intelligence and Soft Computing Research, Vol. 7, No. 1, 2017, pp. 17-31.

[50] Y.G. Yatracos, Rates of convergence of minimum distance estimators and Kolmogorov's entropy. The Annals of Statistics, Vol. 13, 1985, pp. 768-774.

Tomasz Gałkowski is an Assistant Professor at the Institute of Computational Intelligence, Czestochowa University of Technology, Częstochowa, Poland. He graduated from the Electrical Engineering Department of CUT and obtained the M.Sc. in 1980. He received the Ph.D. degree with distinction from Wroclaw University of Technology in 1989. Dr. Gałkowski's research interests lie in the area of pattern recognition and applications, computational intelligence methodologies, data mining, telecommunication, cryptography, and cybersecurity.

Adam Krzyżak received the M.Sc. and Ph.D. degrees in computer engineering from the Wrocław University of Science and Technology, Poland, in 1977 and 1980, respectively, and D.Sc. degree (habilitation) in computer engineering from the Warsaw University of Technology, Poland in 1998. In 2003 he received the Title of Professor from the President of the Republic of Poland. Since 1983, he has been with the Department of Computer Science and Software Engineering, Concordia University, Montreal, Canada, where he is currently a Professor. In 1983, he held an International Scientific Exchange Award in the School of Computer Science, McGill University, Montreal, Canada, in 1991, the Vineberg Memorial Fellowship at the Technion Israel Institute of Technology and, in 1992, Humboldt Research Fellowship at the University of Erlangen-Nurnberg, Germany. He visited the University of California Irvine, Information Systems Laboratory at Stanford University, Riken Frontiers Research Laboratory, Japan, Stuttgart University, Technical University of Berlin, University of Saarlandes and Technical University Darmstadt. He published over 300 papers on neural networks, pattern recognition, nonparametric estimation, image processing, computer vision and control. He has been an associate editor of IEEE Transactions on Neural Networks and IEEE Transactions on Information Theory and is presently an Associate Editor-in-Chief of the Pattern Recognition Journal. He was co-editor of the book Computer Vision and Pattern Recognition (Singapore: World Scientific, 1989) and is a co-author of the book A Distribution-Free Theory of Nonparametric Regression, New York: Springer, 2002. He has been co-chair of the Program Committee of the 10-th IEEE International Conference on Advanced Video and Signal-Based Surveillance 2013 and International Conference on Pattern Recognition and Artificial Intelligence 2018. He has served among others on the program committees of Vision Interface Conference, International Conference on Document Processing and Applications, International Conference on Computer Vision, Pattern Recognition and Image Processing and International Conference on Pattern Recognition. He co-organized a workshop at NIPS'94 Conference and was a session organizer at The World Congress of Nonlinear Analysts in 2000, 2004 and 2008. He is a Fellow of the IEEE.

Zbigniew Filutowicz is an assistant professor at the University of Social Sciences in Łódź. He received the M.Sc. and Ph.D. degrees from Lodz University of Technology in 1973 and 1982, respectively. His research interests include human-computer communication, software engineering and applications of artificial intelligence in computer graphics and medical dialysis.