Journal of Mathematical Imaging and Vision 16: 223–235, 2002

Adaptive Window Size Image De-noising Based on Intersection of Confidence Intervals (ICI) Rule

VLADIMIR KATKOVNIK
Department of Mechatronics, Kwangju Institute of Science and Technology Kwangju 500-712, Republic of Korea

KAREN EGIAZARIAN & JAAKKO ASTOLA
Signal Processing Laboratory, Tampere University of Technology, FIN-33101, Tampere, Finland

Abstract. We describe a novel approach to solve a problem of window size (bandwidth) selection for filtering an image signal given with a noise. The approach is based on the intersection of confidence intervals (*ICI*) rule and gives the algorithm, which is simple to implement and nearly optimal in the point-wise mean squared error risk. The local polynomial approximation (*LPA*) is used in order to derive the 2*D* transforms (filters) and demonstrate the efficiency of the approach. The *ICI* rule gives the adaptive varying window size and enables the algorithm to be spatially adaptive in the sense that its quality is close to that which one could achieve if the smoothness of the estimated signal was known in advance. Optimization of the threshold (design parameter of the *ICI*) is studied. It is shown that the cross-validation adjustment of the threshold significantly improves the algorithm accuracy. In particular, simulation demonstrates that the adaptive transforms with the adjusted threshold parameter perform better than the adaptive wavelet estimators.

Keywords: local adaptive window size transforms, local polynomial approximation, window size selection

1. Introduction

A noise removal (de-noising) is one of the important problems in image processing. Among other approaches to this problem the local polynomial approximation (*LPA*) can be treated as probably one of the most theoretically justified and well studied ones. Originally *LPA* was proposed and developed in statistics for processing scalar and multidimensional noisy data. It is a powerful nonparametric technique which provides estimates in a point-wise manner based on a mean square polynomial fitting in a sliding window e.g. [1, 3, 5, 6, 13, 16]. In terms of the signal and image processing the *LPA* is a flexible tool to design 2*D* transforms with prescribed reproductive properties with respect to polynomial (smooth) components of signals. The invariant and variant optimal window size selection has been studied thoroughly by many authors. These optimal, in particular, varying data-driven window size methods are of special interest for the problems where the piece-wise smooth approximations are the most natural and relevant ones. Some image de-noising problems provide good examples of these cases.

A crucial difference between the nonparametric *LPA* estimates and the more traditional parametric ones, say the polynomial mean squared estimates, is that the latter are formed as unbiased ones while the former are biased on the definition and the reasonable choice of the biasedness controlled by the value of the window size is of importance. It can be emphasized that the problem of the optimal window size selection admits an accurate mathematical formulation in terms of the nonparametric approach, where the optimal window size is defined by a compromise between the bias and the variance of estimation.

Two main ideas are exploited for adaptive (data-driven) window size selection. The first one is based on estimation of the biasedness and the variance of the estimates of the signal with the corresponding optimal window size calculation based on theoretical formulas. However, the bias depends on the derivatives of a given signal. Thus, we need to estimate these derivatives and for this purpose to select auxiliary window sizes. Actually this sort of methods, known as "pilot estimates", is quite complex in implementation and have quite a few design parameters.

The second alternative idea does not have to deal with the bias estimation. This group of methods is based on the quality-of-fit statistics such as the cross-validation, generalized cross-validation, C_p, Akaike criteria, etc., which are applied for direct optimization of the accuracy.

A number of publications concerning the window size selection problem is very large and growing quickly. A review on the field or even its brief analysis is far beyond of the goal of this paper. Here we give only few references illustrating the basic progress in different directions.

A successful implementation of the first approach based on the pilot estimates has been reported by several authors. An automatic local window size selector with estimation of the higher order derivatives of $y(x)$, which are plugged into the local risk expression, was developed in [3]. The empirical-bias window size selection [19] uses the estimates at several window sizes in order to approximate the bias, and results in quite an efficient adaptive smoother for estimation of the function and its derivative. The similar ideas have been exploited in the adaptive smoothers described in [20].

Most of publications concerning the second approach are related to a data-based global (constant) window size selection e.g. [6, 7, 13]. The linear *LPA* with the varying window size found by minimization the so-called "pseudo-mean squared error" is considered in [18]. The target point is left out of the averaging in the pseudo-mean squared error what differs this method from the standard mean square methods. It is reported that the proposed pseudo-mean squared error works better then the local cross-validation [18].

This paper is based on a quite recent new development. The intersection of confidence intervals (ICI) rule originally was proposed and developed in [4] and [8] for de-noising of 1*D* observations and shown to be quite efficient in particular for many distinct applications e.g. [10, 12, 14]. First results on applications of the ICI rule to image de-noising have been reported in [9, 15].

This paper presents a systematic development of the *ICI* rule for image de-noising including basic ideas, algorithms, theoretical performance analysis and simulation results. We introduce adaptive 2*D* transforms for smoothing an image intensity function given with an additive noise. These transforms are able to produce a piece-wise smooth surface with a small number of discontinuities in the intensity function or its derivatives. This allows certain desirable features of images such as jumps or instantaneous slope changes to be preserved.

The further paper is organized as follows. In section 2 the *LPA* technique is described in details with theoretical results concerning the accuracy analysis. The idea and algorithm implementation of the *ICI* window size selection is described in section 3 as well as the data-driven adjustment of the threshold parameter and the multiple window *LPA*. Simulation, given in section 4 illustrates a good performance of the proposed methods as well as some modifications of the basic algorithms.

2. Local Polynomial Approximation

2.1. Transforms Based on LPA

We start with a discussion of 2*D* linear transforms derived on the base of the *LPA* method.

Suppose that we are given by noisy observations of the image intensity $y(x)$, $y \in R^1$, $x \in R^2$, on the regular or irregular grid of argument values $x(s) = (x_1(s), x_2(s))$ being two dimensional vectors with components $x_1(s)$ and $x_2(s)$ and the parameter s indicating corresponding $s-th$ pixel of the image.

Then, the noisy data can be given as:

$$z(x) = y(x) + \varepsilon(x) \tag{1}$$

where $\varepsilon(x)$ are independent and identically distributed (i.i.d.) random Gaussian errors, $E[\varepsilon(x)] = 0$, $E[\varepsilon^2(x)] = \sigma^2$. It is assumed that $y(x)$ belongs to a nonparametric class of piece-wise continuous $r-$differentiable functions

$$\mathcal{F}_r = \{|D^r_{r_1,r_2} y(x)| \le L_r(x) \le \bar{L}_r, \forall\, r_1 + r_2 = r\}, \tag{2}$$

where $D^{r_1+r_2}_{r_1,r_2} = \frac{\partial^{r_1+r_2}}{\partial^{r_1}x_1\partial^{r_2}x_2}$ is a differentiation operator and $\bar{L}_r$ is a finite constant.

Our goal is to estimate $y(x)$ depending on the observations $\{z(x)\}$ with the point-wise mean squared error (*MSE*) risk which is as small as possible.

The *LPA* of $y(x)$ is applied in the following form. First, a part of the truncated Taylor series is used in order to approximate the varying intensity $y(x)$, and second, this expansion is exploited locally in a comparatively small area. In fact, the local expansion is applied in order to calculate the estimate for a single "central" pixel only. For the next pixel the calculations should be repeated. This point-wise procedure determines a nonparametric character of estimation.

The following criteria function is applied in the *LPA* e.g. [1, 3, 6, 13, 16]:

$$J_h(x) = \sum_s w_h(x(s)-x)\,(z(x(s)) - C^T\phi(x(s)-x))^2 \quad (3)$$

$$\phi(x) = (\phi_1(x), \phi_2(x), ..., \phi_M(x))^T,$$
$$C = (C_1, C_2, ..., C_M)^T, x(s) = (x_1(s), x_2(s)), \quad (4)$$

where $\phi(x)$ is a set of linear independent $2D$ polynomials of the powers from 0 to m with $\phi_1 = 1$. A total number of these polynomials is equal to $M = C_2^{2+m} = \frac{(2+m)!}{2\cdot m!} = \frac{(2+m)(1+m)}{2}$. The term "coordinate function" is sometimes used for ϕ_k. The window $w_h(x) = w(x/h)/h$ formalizes the localization of fitting with respect to the centre x, while the scale parameter $h > 0$ determines the size of the window. The windowing weight $w_h(x) = w(x/h)/h$ is a function satisfying the conventional properties: $w(x) \geq 0$, $w(0) = \max_x w(x)$, $\int\int w(x_1, x_2)\,dx_1dx_2$=1.

A multiplicative window $w(x) = w_1(x_1)w_2(x_2)$, where $w_1(x_1)$ and $w_2(x_2)$ are functions of scalar arguments, is used in many applications. If the window is rectangular all observations enter in the criteria function with equal weights . Nonrectangular windows such as triangular, quadratic, Epanechnikov and so on [1, 3, 6, 13, 16] usually prescribe higher weights to observations which are closer to the centre x.

The point-wise *LPA* insures the reproduction properties of the estimate with respect to the polynomial components of $y(x)$. But it should be emphasized that the *LPA* estimate of $y(x)$ is not a polynomial function. This is a principal difference between the nonparametric *LPA* and the corresponding parametric models. For $m = 2$ a full set of the linear independent polynomial is of the form

$$\phi_1 = 1, \text{for } m = 0;$$
$$\phi_2 = x_1, \phi_3 = x_2, \text{for } m = 1;$$
$$\phi_4 = x_1x_2, \phi_5 = x_1^2/2, \phi_6 = x_2^2/2, \text{for } m = 2; \quad (5)$$

with $M = \frac{(2+m)!}{2\cdot m!} = 6$.

Minimizing $J_h(x)$ with respect to C,

$$\hat{C}(x,h) = \arg\min_{C\in R^M} J_h, \quad (6)$$

gives $\hat{y}(x) \stackrel{\Delta}{=} \hat{C}_1(x,h)$ as an estimate of $y(x)$, and $\hat{C}_l(x,h)$, $l = 2, ..., M$, as estimates of the derivatives of $y(x)$. For the polynomials (5) these estimated derivatives are $D^1_{1,0}y(x)$, $D^1_{0,1}y(x)$, $D^2_{1,1}y(x)$, $D^2_{0,2}y(x)$, $D^2_{2,0}y(x)$ respectively. Recall that the first and second order derivatives $(D^{k_1+k_2}_{k_1,k_2}y(x), 1 \leq k_1 + k_2 \leq 2)$ are used as a tool for image segmentation and enhancement. However, in this paper we concern with the estimation of the intensity function only.

These estimates can be represented in the form of a linear transform (filter)

$$\hat{y}(x,h) = \sum_s g_1(x, x(s), h)z(x(s)), \quad (7)$$

where g_1 is a first element of the vector g given by the equation

$$g = \Phi^{-1}w_h(x(s)-x)\phi(x(s)-x),$$
$$\Phi = \sum_s w_h(x(s)-x)\phi(x(s)-x)\phi^T(x(s)-x). \quad (8)$$

It can be verified that for any polynomial $y(x)$ of the power m the estimate (7) is accurate. In particular, for the polynomials (5) it means that

$$\sum_s g_1(x, x(s), h)\phi_k(x(s)) = \phi_k(x),$$
$$\text{for } k = 1, ..., M, \forall x. \quad (9)$$

It shows that the transform with the weight g_1 has an accurate reproductive properties for the $2D$ polynomial components of the intensity up to the power $m = 2$.

The linear transform (7)-(8) can be applied to data given on any regular or irregular grids, in particular, to data with lost observations and for data interpolation problem when x does not belong to the grid $\{x(s)\}$. It

is assumed in the formulas (7)-(8) that the summation is always performed within boundaries of the image frame.

For the regular infinite grid and x belonging to this grid the linear estimators (7)-(8) can be written as a homogeneous (stationary) transform:

$$\hat{y}(x,h) = \sum_s g_1(x(s),h)\, z(x+x(s)), \tag{10}$$

where g_1 is a first element of the vector g

$$\begin{aligned} g &= \Phi^{-1} w_h(x(s))\, \phi(x(s)), \\ \Phi &= \sum_s w_h(x(s))\, \phi(x(s))\, \phi^T(x(s)). \end{aligned} \tag{11}$$

Then, the equations (9) for the polynomials (5) have a form of moment restrictions on the weight function g_1

$$\begin{aligned} &\sum_s g_1(x(s),h) = 1, \sum_s g_1(x(s),h) x_1(s) = 0, \\ &\sum_s g_1(x(s),h) x_2(s) = 0, \\ &\sum_s g_1(x(s),h) x_1(s) x_2(s) = 0, \\ &\sum_s g_1(x(s),h) x_1^2(s) = 0, \\ &\sum_s g_1(x(s),h) x_2^2(s) = 0. \end{aligned} \tag{12}$$

An important difference between the estimates (7)-(8) and (10)-(11) is that the latter assumes that a number of observations in the estimate does not depend on the location of x in the image. As a result the function $z(x)$ should be defined beyond the boundaries of the image frame. Naturally, the accurate polynomial fitting is not fulfilled in this case for pixels in a neighborhood of the boundaries of the frame. The estimate in the form (7)-(8) is free from this boundary effect.

Actually, the design of the weights g_1 by the formulas (7)-(8) or (10)-(11) is quite of a general nature. Say, we can start from the zero order approximation with the only one coordinate function $\phi_1 = 1$. In this case the fitter gives an approximation by a constant into the sliding window. It is able to guaranty the accurate unbiased reproduction only for the constant component of the intensity. Further, the linear functions $\phi_2 = x_1$ and $\phi_3 = x_2$ can be added to $\phi_1 = 1$ as elements of the vector $\phi = (\phi_1,\ \phi_2,\ \phi_3)^T$. The corresponding transform is able to obtain the accurate reproduction of constant and linear on x_1 and x_2 components of the intensity. As a next step, the cross term $\phi_4 = x_1 x_2$ can be involved into the *LPA* by including in $\phi = (\phi_1,\ \phi_2,\ \phi_3,\ \phi_4)^T$. Further, one or both of the quadratic functions $\phi_5 = x_1^2/2$, $\phi_6 = x_2^2/2$ can be used in the *LPA*. Thus, it is not necessary to use a full set of the coordinate functions ϕ_k of some particular power m. The considered design of the transform can be used in order to obtain desirable properties with respect to any components of the intensity function. Naturally, non-polynomial coordinate functions can be applied in the local approximation in a straightforward manner.

The linear estimators (7)-(8) and (10)-(11) have a very long prehistory e.g. [1, 3, 5, 6, 13, 16]. They are a very popular tool in statistics and signal processing with application to a wide variety of the fields for smoothing, filtering, differentiation, interpolation and extrapolation.

2.2. *Accuracy of the LPA*

It is well known that window size selection is a crucial point of the efficiency of the local estimators. When h is relatively small, the *LPA* gives a good approximation of $y(x)$ but then fewer data are used and the estimates are more variable and sensitive with respect to the noise. The best choice of h involves a trade-off between the bias and variance, which depends on the order of the derivatives being involved in the *LPA*, a sample period, the noise variance, and values of the derivatives of $y(x)$ beyond the order used in the *LPA*.

The estimation error can be represented as follows

$$\begin{aligned} e(x,h) &= y(x) - \hat{y}(x,h) \\ &= y(x) - \sum_s g_1(x,x(s),h) z(x(s)) \\ &= E(e(x,h)) + e^0(x,h), \end{aligned} \tag{13}$$

where $E(e(x,h)) = y(x) - \sum_s g_1(x,x(s),h)\, y(x(s)) = \sum_s g_1(x,x(s),h)\, [y(x) - y(x(s))]$ and $e^0(x,h) = -\sum_s g_1(x,x(s),h)\, \varepsilon(x(s))$ are the bias and random component of the estimation error.

The Taylor series of the power r with the residual terms in Lagrangian's form is used for the difference $y(x) - y(x(s))$:

$$y(x)-y(x(s))=S_1(s)+S_2(s);$$
$$S_1(s)=\sum_{k_1+k_2=1}^{r-1}\frac{1}{k_1!k_2!}[x_1-x_1(s)]^{k_1}[x_2-x_2(s)]^{k_2} D^{k_1+k_2}_{k_1,k_2}y(x);$$
$$S_2(s)=\sum_{k_1+k_2=r}\frac{1}{k_1!k_2!}[x_1-x_1(s)]^{k_1}[x_2-x_2(s)]^{k_2} D^{r}_{k_1,k_2}y(\lambda_s x(s)+(1-\lambda_s)x),$$
$$0\le\lambda_s\le 1. \quad (14)$$

Assume that ϕ in (7)-(8) is a full set of $2D$ polynomials of the power m and $r=m+1$. Then, according to (9),

$$\sum_s g_1(x,x(s),h)S_1(s)=0$$

and the bias is defined by the $(m+1)-th$ derivatives $D^{m+1}_{k_1,k_2}y$.

Substituting (14) into the formula for the bias-error, we obtain

$$|E(e(x,h))|\le\bar{L}_{m+1}\sum_{k_1+k_2=m+1}\frac{1}{k_1!k_2!}\sum_s|g_1(x,x(s),h)||x_1-x_1(s)|^{k_1}|x_2-x_2(s)|^{k_2},$$

where according to (2) we used that $|D^{m+1}_{k_1,k_2}y|\le\bar{L}_{m+1}$.

The variance of the random components is given by the equation

$$var(x,h)=E(e^0(x,h)^2)=\sigma^2\sum_s(g_1(x,x(s),h))^2. \quad (15)$$

In order to derive the formulas which provide a clear dependence of the accuracy on the window width parameter h we assume that the sampling period Δ is small, $\Delta\to 0$, and the $2D$ sampling grid is regular. The sums in the all above formulas can be transformed into integrals, and, after some manipulations, we arrive to the following expressions:

$$\hat{y}(x,h)=\int\int g_1(u)y(x+hu)\,du_1du_2,$$
$$g=\Phi^{-1}w(u)\phi(u),$$
$$\Phi=\int\int w(u)\phi(u)^T(u)\,du_1du_2,$$
$$g=(g_1,...,g_M)^T,u=(u_1,u_2). \quad (16)$$

Then, the formulas for the bias and variance can be given in the explicit analytical form

$$|E(e(x,h))|\le h^{m+1}L_{m+1}(x)A,$$
$$A=\sum_{k_1+k_2=m+1}\frac{1}{k_1!k_2!}\int\int|g_1(u)||u_1|^{k_1}|u_2|^{k_2}du_1du_2,$$
$$var(x,h)\simeq\frac{\Delta^2\sigma^2}{h^2}B,B=\int\int|g_1(u)|^2du_1du_2. \quad (17)$$

Thus, the point-wise mean squared risk $r(x,h)$ in asymptotic with a small Δ can be represented as follows

$$r(x,h)\triangleq E(e(x,h))^2\le\bar{\omega}^2(x,h)+\frac{\Delta^2\sigma^2}{h^2}B\triangleq\bar{r}(x,h),$$
$$\bar{\omega}(x,h)=h^{m+1}L_{m+1}(x)A, \quad (18)$$

where $\bar{\omega}(x,h)$ denotes the upper bound of the bias.

Minimizing on h the upper bound $\bar{r}(x,h)$ of the mean squared risk gives for the ideal values of the window size and the risk upper bound:

$$h^*(x)=\left(\Delta^2\frac{\sigma^2B}{A^2(L_{m+1}(x))^2}\gamma^2\right)^{1/(2m+4)},\gamma^2=\frac{1}{m+1} \quad (19)$$

and

$$\bar{r}^*(x)=\bar{r}(x,h^*(x))=var^*(x)(1+\gamma^2),$$
$$var^*(x)=var(e^0(x,h^*(x))),\gamma=\bar{\omega}^*(x)/std^*(x),$$
$$\bar{\omega}^*(x)=\bar{\omega}(x,h^*(x)),std^*(x)=\sqrt{var^*(x)}, \quad (20)$$

where the constant γ which is not depending on x, shows a proportion between the upper bound of the bias and standard deviation of the estimation error at the ideal window size $h^*(x)$.

The formulas (19) and (20) demonstrate that the ideal window size $h^*(x)$ depends on the $(m+1)-th$ derivatives of $y(x)$ and the ideal variance-bias trade-off is achieved when the ratio between the bias and standard deviation $\omega^*(x)/std^*(x)$ is equal to γ. It can be seen that

$$\bar{\omega}(x,h)\begin{cases}<\gamma\cdot std(x,h)\ if\ h<h^*\\ >\gamma\cdot std(x,h)\ if\ h>h^*\end{cases} \quad (21)$$

In what follows this inequality is used in order to test the hypotheses: $h\lesssim h^*$.

3. Adaptive Window Size Selection

3.1. The idea of the ICI

The estimation error of the *LPA* can be represented in the form

$$|e(x,h)| = |y(x) - \hat{y}(x,h)| \le \bar{\omega}(x,h) + \left|e^0(x,h)\right|, \quad (22)$$

where $\bar{\omega}(x,h)$ is the upper bound of the estimation bias and $e^0(x,h)$ is a random error with the probability density $N(0, std^2(x,h))$. Then $\left|e^0(x,h)\right| \le \chi_{1-\alpha/2} \cdot std(x,h)$ holds with the probability $p = 1-\alpha$, where $\chi_{1-\alpha/2}$ is $(1-\alpha/2) - th$ quantile of the standard Gaussian distribution, and with the same probability

$$|e(x,h)| \le \bar{\omega}(x,h) + \chi_{1-\alpha/2} std(x,h). \quad (23)$$

It follows from (21) that the inequality (23) can be weakened to

$$|e(x,h)| \le (\gamma + \chi_{1-\alpha/2})\, std(x,h). \quad (24)$$

Now let us introduce a finite set of window size:

$$H = \{h_1 < h_2 < .... < h_J\},$$

starting with a quite small h_1, and, according to (24), determine a sequence of the confidence intervals $\mathcal{D}(j)$ of the biased estimates as follows

$$\mathcal{D}(j) = [\hat{y}(x,h_j) - \Gamma \cdot std(x,h_j), \hat{y}(x,h_j) + \Gamma \cdot std(x,h_j)], \quad (25)$$

where

$$\Gamma = \frac{1}{\sqrt{m+1}} + \chi_{1-\alpha/2} \quad (26)$$

is the threshold of the confidence interval.

Then for $h = h_j$ (24) is of the form

$$y(x) \in \mathcal{D}(j), \quad (27)$$

and we can conclude from (23) and (24) that while $h_j < h^*$ holds for $h = h_j$, $1 \le j \le i$, all of the intervals $\mathcal{D}(j)$, $1 \le j \le i$, have a point in common, namely, $y(x)$.

In the opposite case, when the intersection of the confidence intervals is empty it indicates that $h_j > h^*$. Thus, the intersection of the confidence intervals can be used in order to verify the inequalities (21).

The following is the *ICI* statistic, which is used in order to test the very existence of this common point and in order to obtain the adaptive window size value:

Consider the intersection of the intervals $\mathcal{D}(j)$, $1 \le j \le i$, *with increasing i, and let* i^+ *be the largest of those i for which the intervals* $\mathcal{D}(j)$, $1 \le j \le i$, *have a point in common. This* i^+ *defines the adaptive window size and the adaptive LPA estimate as follows*

$$\hat{y}^+(x) = \hat{y}(x, h^+(x)), h^+(x) = h_{i^+}. \quad (28)$$

3.2. Algorithm

The following algorithm implements the procedure (28). Determine the sequence of the upper and lower bounds of the confidence intervals $\mathcal{D}(i)$ as follows

$$\begin{aligned} \mathcal{D}(i) &= [L_i, U_i], \\ U_i &= \hat{y}(x,h_i) + \Gamma \cdot std(x,h_i), \\ L_i &= \hat{y}(x,h_i) - \Gamma \cdot std(x,h_i), \end{aligned} \quad (29)$$

where Γ is given by (26).

Let

$$\begin{aligned} &\bar{L}_i + 1 = \max[\bar{L}_i, L_{i+1}], \underline{U}_{i+1} = \min[\underline{U}_i, U_{i+1}], \\ &i = 1,2,...,J, \bar{L}_1 = L_1, \underline{U}_1 = U_1 \end{aligned} \quad (30)$$

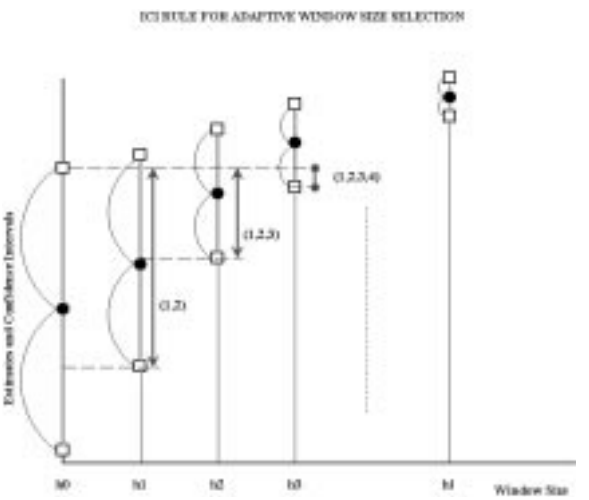

Figure 1. Graphical illustration of ICI rule.

then the adaptive window length h_i^+ is the largest i when

$$\bar{L}_i \le \underline{U}_i \tag{31}$$

is still satisfied. This i^+ is the largest of those i for which the confidence intervals $\mathcal{D}(i)$ have a point in common as it is discussed above.

We wish to emphasize that this window size *ICI* selection procedure requires a knowledge of the estimate and its variance only.

The *ICI* rule is graphically illustrated in Fig. 1, where the vertical lines with arrows show the successive intersections of the confidence intervals (1, 2), (1, 2, 3), and (1, 2, 3, 4). Assuming that the intersection with the forth confidence interval (corresponding $h = h_4$) is empty, we obtain the "optimal" adaptive window size $h^+ = h_3$.

3.3. *Adjustment of the Threshold* Γ

The threshold parameter Γ in (29) plays a crucial role in the performance of the algorithm. Too large or too small Γ results in oversmoothing and undersmoothing data.

Let us present some figures for Γ following from the theoretical analysis. Assuming $\alpha = 0.05$ or 0.01, then $\chi_{1-\alpha/2} = 2$ or 3 respectively, we obtain for the threshold:

$$\Gamma = \begin{cases} 3.0, \text{ for } p = 0.05, \\ 4.0, \text{ for } p = 0.01, \end{cases} \qquad \text{for } m = 0,$$

$$\Gamma = \begin{cases} 2.7, \text{ for } p = 0.05, \\ 3.7, \text{ for } p = 0.01. \end{cases} \qquad \text{for } m = 1,$$

Remind that the formula for γ (19) determining the optimal threshold Γ is obtained for the asymptotic as $\Delta \to 0$ and provided that the intensity $y(x)$ is smooth enough in the neighborhood of x, i.e. the first and second order derivatives exist respectively when the zero order ($m = 0$) or the first order ($m = 1$) *LPA* is applied. In practice, the smoothness of $y(x)$ can not be guaranteed for every x and Δ can be not small. Another uncertain point concerns the confidence level α and quantile $\chi_{1-\alpha/2}$ which are used in calculations of the threshold Γ. Thus, we may conclude, that there are ambiguities, which influence a selection of the threshold and these ambiguities cannot be resolved in terms of the theoretical analysis only. However, the threshold Γ is a natural invariant design parameter of the algorithm, which can be used in order to refine the algorithm and to adjust it to the available observations.

We produced a number of Monte-Carlo simulation experiments in order to verify a role of Γ. In particular, the *MSE* of de-noising is minimized on Γ in every Monte-Carlo simulation run. The optimal values of Γ found in this way are random but have very small variations. Actually, it means that these optimal values of Γ depend on statistical properties of the noise but not particular samples used in the Monte-Carlo runs. These optimal Γ are quite robust with respect to random noise components of observations.

The cross-validation (*CV*) is one of the popular tools developed in quality-of-fit statistics for model selection and adjustment e.g. [7]. For the linear estimator in the form (7) the *CV* loss function can be represented as a weighted sum of squared residuals e.g. [8]:

$$I_{CV} = \sum_s \left(\frac{z_s - \hat{y}(x(s),\, h^+(x(s)))}{1 - g_1(x(s),\, x(s),\, h^+(x(s)))} \right)^2. \tag{32}$$

Thus, the procedure (29)-(31) is assumed to be repeated for every $\Gamma \in G$, $G = \{\Gamma_1, \Gamma_2, ...\Gamma_{N_G}\}$, and

$$\hat{\Gamma} = \arg\min_{\Gamma \in G} I_{CV} \tag{33}$$

gives the adjusted threshold parameter value.

The cross-validation in the form (32) presents quite a reasonable and efficient selector for Γ. Our attempts to use instead of the cross-validation another quality-of-fit statistics, in particular the C_P, Akaike criteria and its modifications (see e.g. [7]), which are different from I_{CV} only by the used weights of the residuals, have not shown an improvement in accuracy.

The adjusted adaptive *LPA* estimation consists of the following basic steps:

1. Set $\Gamma = \Gamma_l$, $l = 1, 2, ..., N_G$ and $x = x(s), s = 1, 2, .., N$.
2. For $h = h_i$, $i = 1, ..., J$, calculate the estimates $\hat{y}(x(s), h)$, the adaptive window size $h^+(x(s))$ and the estimate $\hat{y}(x(s), h^+(x(s)))$.
4. Repeat Step 2 for all $x(s)$, $s = 1, 2, .., N$, and Γ_l, $l = 1, 2, ..., N_G$.
5. Find $\hat{\Gamma}$ from (33) and select estimates $\hat{y}(x(s), h^+(x(s))$ corresponding to $\hat{\Gamma}$ as the final ones.

The standard deviation σ used in $std(x,h)$ is estimated by

$$\hat{\sigma} = \{\text{median}(|z_s - z_{s-1}| : \\ s = 2,..,N)\}/(\sqrt{2}\cdot 0.6745). \quad (34)$$

The average $\frac{1}{N-1}\sum_{n=2}^{N}(z_s - z_{s-1})^2$ could also be applied as an estimate of σ^2. However, we prefer a median (34) as a robust estimate.

3.4. Multiple Window Estimation

Different ideas can be used for a design of the varying window for processing of $2D$ image signals. The simplest and standard one assumes that a symmetric square window is applied for every pixel and the size of the window is the only varying parameter to be found.

A more complex approach assumes that the varying window is composed from a number of separate segments, say from four quadrants shown in Figure 2. The centre of the window is the initial point of the Cartesian coordinate system $(0,0)$. Each segment is a square covering a part of the corresponding quadrant. It is assumed that this initial point $(0,0)$ is the centre of the *LPA* estimate for each square segment. The sizes of these squares are the parameters of the combined window. The *ICI* rule is used for independent selection of the sizes of these separate four windows. There are a number of ways to fuse estimates obtained for the separate window segments into the one final estimate.

Some of our simulation results presented in this paper are obtained for the following final estimate:

$$\hat{y}(x) = \sum_{j=[1,2,3,4]} \lambda_j \hat{y}_j(x, h_j^+(x)), \\ \lambda_j = \frac{std_j^{-2}}{std^{-2}}, std^2 = \sum_{j=[1,2,3,4]} std_j^2, \quad (35)$$

where $\hat{y}_j(x, h_j^+(x))$ are the estimates with the *ICI* rule adaptive window size, $j = [1, 2, 3, 4]$, obtained respectively for the windows 1, 2, 3, 4 in Figure 2. Further λ_j and std_j are the weights and the standard deviations of these estimates $\hat{y}_j(x, h_j^+(x))$. In the estimate (35) we use a linear fusing of the estimates with the inverse standard deviations of the estimates as weights. Similar multiple combined window estimates have been applied in [4] and [8] for $1D$ function estimation.

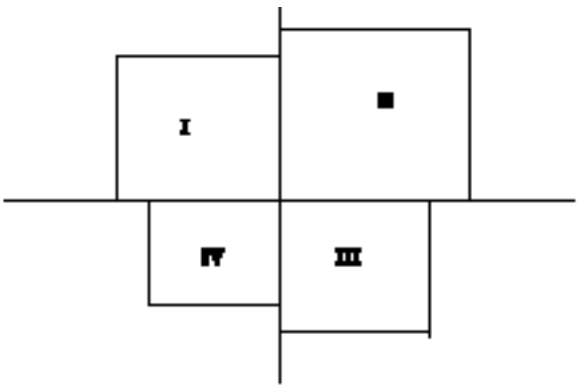

Figure 2. Four quadrant windows 1, 2, 3 and 4 used for directional window size selection by the ICI rule.

4. Algorithms and Simulation Results

The *ICI* rule for window size selection and the multiple window estimates introduced in Section 3 define a basic algorithm developed for noise reduction. A number of modifications of this algorithm has been developed and studied. These modifications use different methods in order to form the combined window, special corrections of the adaptive window size given by the *ICI* rule and the different estimation methods applied to the data in these varying windows. In particular, an adaptive size and shape window growing pixel-by-pixel using the *ICI* rule is proposed in [2]. In this paper the *ICI* rule is used for the pointwise varying window size segmentation of the image only, while the orthogonal transforms estimators (e.g. wavelet and discrete cosine transform (*DCT*)) different from the *LPA* are applied for de-noising in this varying size segments.

The median filters equipped with the *ICI* for the varying window size selection are reported in [11, 14].

It was noticed that the *ICI* adaptive window sizes, in particular for small Γ, can be corrupted by spikes which erroneously isolate small values to the window sizes [8]. The preliminary filtering of $h_j^+(x)$ considered as a function of x, say by a simple median filter, is able to improve the quality of de-noising.

In this section we present simulation results which illustrate the efficiency of the *ICI* rule and give an insight into behavior of the estimates. We consider also some modified versions of the basic algorithm.

1. Let us start from a simple binary image. Figure 3 shows a true image, noisy image and de-noised image. The basic LPA transform with $m = 0$ and $\Gamma = 4.0$ is used. The adaptive window sizes for windows 1, 2, 3 and 4 are given in Figure 4. Small and large window sizes are shown there by black and white, respectively. Isolated black points in Figure 4 are spikes corresponding to random small window sizes given by the *ICI* rule. It deserves to be mentioned that these isolated spikes have different locations into four different windows and do not influence the final de-noised image shown in Figure 3. Actually these spikes can be eliminated by increasing the value of Γ. However, it results in increasing of the value of the root mean squared error (*RMSE*).

Presenting Figure 4 we emphasize that the obtained window sizes actually correspond to the intuitively clear behavior of the varying window size relevant to the smoothing of the data if the true image is known. Thus the window sizes delineate the true image of the square and the variations of the window sizes provides a shadowing of the image from different sides of the image in full agreement with the directional windows used for smoothing (see Figure 2).

2. Now let us demonstrate a different algorithm of using the same ICI rule. The developed algorithm comprises of the following two parts. The first part is applied for a point-wise image segmentation. This segmentation assumes that the *LPA* with *ICI* rule is used for every pixel in order to find the adaptive sizes of four directional rectangular windows as shown in Figure 2. As a result, every pixel can be an entry of many different estimates obtained for adaptive varying size windows with different centers. The second part of the algorithm assumes that the *DCT* transform filtering [2] is applied for every of these adaptive size windows. All obtained estimates are accumulated in a buffer and averaged in order to produce the final estimate for every pixel.

Experiments were performed on the test image "Cameraman" (8 bit gray-scale 256×256 image) corrupted by different types of noise. The results are compared with the wavelet transform based (Haar, Symmlet, Coiflet, Translation Invariant [17]) and Wiener filters. The new algorithm showed a valuable signal-to-noise ration (SNR) improvement (more than 4-5 dB) for most of the cases. Some illustrative images are given in Figure 5. Figure 5a,b show the original and noisy image, while the DCT estimate described above is given in Figure 5d. The *RMSE* values show a valuable original noise reduction. The visual quality is quite acceptable for this level of the noise. In Figure 5c we show as an intermediate results the filtering obtained from the zero order *LPA* (sample averaging). The estimates obtained for four adaptive varying windows are averaged with the weights reciprocal to the variances of these

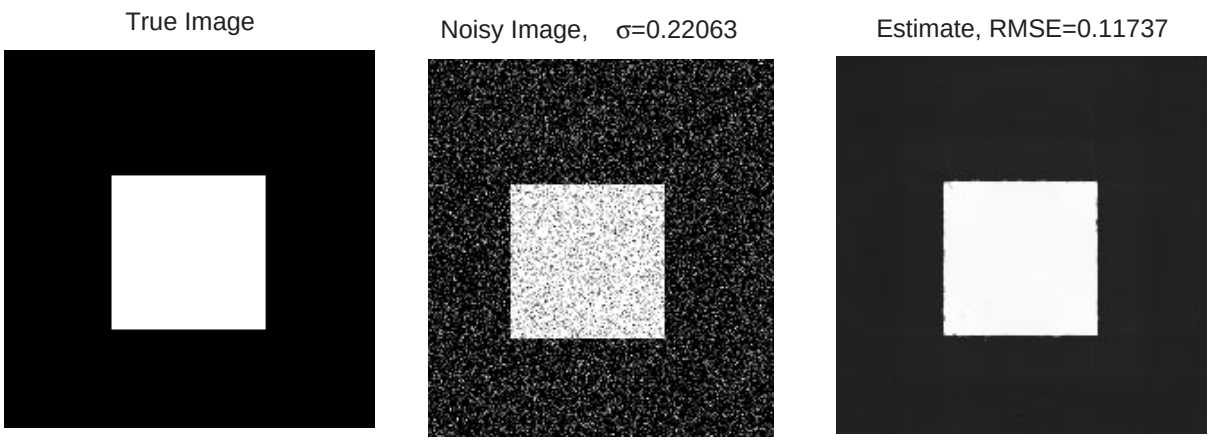

Figure 3. True image, noisy image and denoised image. The basic LPA filter with $m = 0$ and $\Gamma = 4.0$.

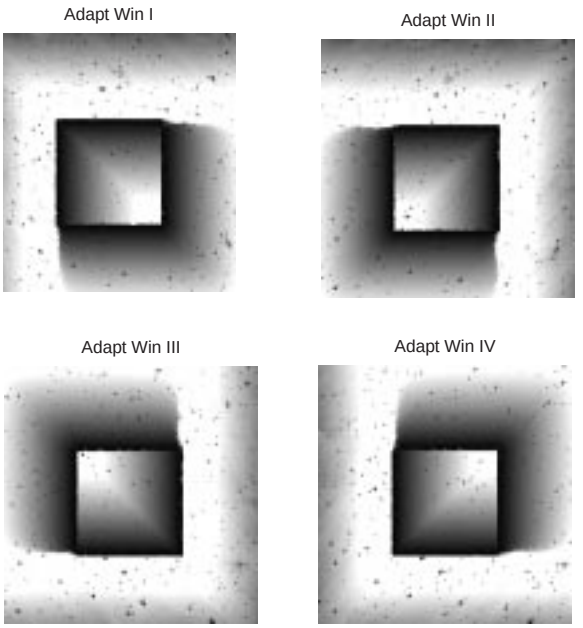

Figure 4. Adaptive window sizes for windows I, II, III and IV. The *LPA* filter with $m = 0$ and $\Gamma = 4.0$. Black and white correspond to small and large window sizes respectively. Isolated black points are spikes in the window size corresponding to random small window sizes by the *ICI* rule.

Figure 5. a) True image, b) Noisy image, c) *LPA* denoising, d) *DCT* denoising with *ICI* adpative window sizes.

estimates. Figure 6 shows the varying adaptive window sizes obtained respectively for the windows I, II, III and IV (Figure 2, $h_k = 2^k$, $k = 0, 1, ..., 6$). Here black and white areas correspond, respectively, to small and large window sizes. Again, the adaptive window sizes delineate contours of the image and demonstrate a very reasonable performance of the *ICI* rule as a window size selector.

Let us discuss on a role of the threshold Γ. Simulation with

$$\Gamma = 1.5, 2.0, 2.5, 3.0, 3.5, 4.0 \qquad (36)$$

was produced with using *DCT* transform filtering and *ICI* window size selection as it is described in Section III. Resulting RMSEs are obtained

$$\begin{aligned} RMSE \;=\; & 0.0806, 0.0605, 0.0706, 0.0811, \\ & 0.0954, 0.1102, \end{aligned} \qquad (37)$$

respectively. Thus, the best performance with $RMSE = 0.0605$ was achieved for $\Gamma = 2.0$. It is the ideal result as it assumes that the true image, used for *RMSE* calculation, was known. Comparison of *RMSE* values shows that the improvement up to 5 dB can be obtained by Γ optimization.

The image presented in Figure 5d is given for the value of the threshold $\Gamma = 2.5$ obtained by the *CV* adjustment of Γ as it is described Section 3. The grid (36) was used in the optimization problem (33). This result is quite close to the optimal $\Gamma = 2.0$. It is well known that an improvement of qualitative criteria, such as *RMSE*, does not guaranty a visual improvement of images. However, the simulation confirms that in terms of this sort of criteria the *CV* can be applied for the adjustment of the de-noising algorithms with varying data-driven window sizes.

3. Here we consider the *DCT* transform filtering equipped with varying adaptive window size. Thus, we apply the *DCT* transform for image de-noising instead of the *LPA* transform. Further, we apply it in the *ICI* rule for the varying window size selection. The latter is done by using in (25) the standard deviation $std(x, h_j)$ of the *DCT* transform. More details on this algorithm as well as its statistical justification can be found in [2].

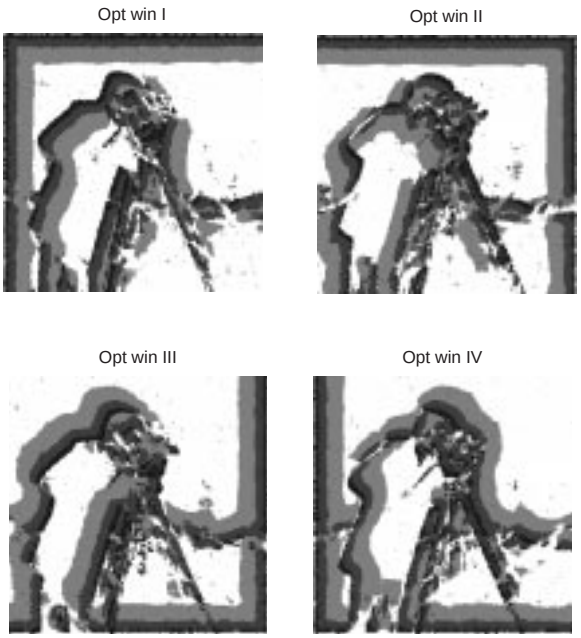

Figure 6. Adaptive window sizes obtained by *ICI* with $\Gamma = 2.5$.

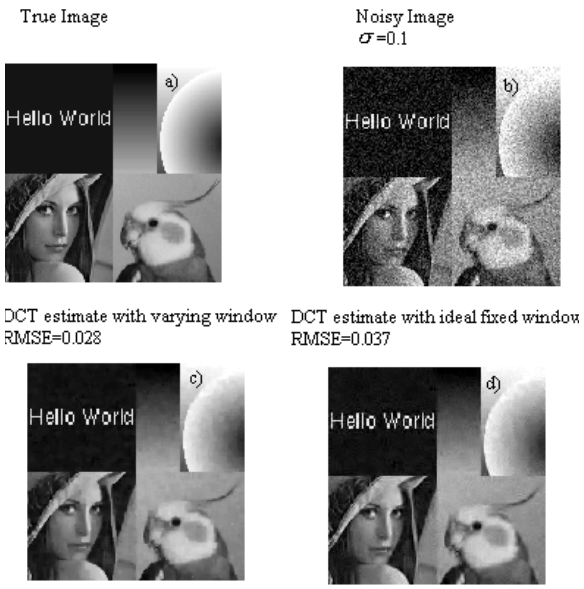

Figure 7. True "Montage" image; noisy image; the estimate with the varying adaptive window size ($\Gamma = 2.0$); the estimate with the fixed ideal size window.

Results of the local adaptive DCT filtering as well as the filtering with the ideal fixed window size *DCT* transform (5×5) are presented in Figure 7. The "Montage" image, composed from different types of subimages, is used in these experiments. RMSE of the filtered signal is equal to 0.028 and 0.037 for the adaptive varying and ideal fixed window size filters, respectively. The filter with the varying window size yields better resolution of details as well as less value of *RMSE*.

Table 2 below presents accuracy results obtained for this montage image by different filters. We compare the local DCT filters considered above versus the Wiener filter and different wavelet based filters. All accuracy results are in favor of the local DCT with adaptive varying window size.

Table 2. Comparative results

Used Filter	RMSE	MAE
Local DCT with adaptive transform size	0.0281	0.0191
Local DCT with ideal transform size	0.0372	0.0211
Wiener Filter(5x5)	0.0408	0.0274
Wavelet Package Haar	0.0585	0.0444
Wavelet Package 8	0.0464	0.0310
Wavelet PO Haar (4 levels)	0.0567	0.0411
Wavelet PO 8 (4 levels)	0.0693	0.0485
Wavelet TI Haar (5 levels)	0.0317	0.0205
Wavelet TI 8 (5 levels)	0.0365	0.0261

RMSE=Root Mean Square Error,
MAE=Mean Absolute Error,
PO= Periodic, TI= Translation invariant

5. Conclusion

A novel approach to solve a problem of varying adaptive window size selection for filtering a noisy image is presented. The *LPA* is used in order to demonstrate the efficiency of the approach, while a possible development to another linear and nonlinear filters (transforms) can be given. The algorithm is simple to implement and requires calculation of the estimates and their standard deviations for a set of the window size values. The adaptive transform is built as J parallel filters, which are different only by the window size $h_j, j = 1, 2, ..., J$, and the selector, which determine the best window size $h^+(x(s))$ and the corresponding estimate $\hat{y}(x(s), h^+(x(s)))$ for every pixel $x(s)$. This selector uses the *ICI* statistic. In can be proved in a similar way as it is done in [4] for $1D$ regression de-noising, that the adaptive algorithm is

nearly optimal in the point-wise risk for estimating the signal and its derivatives.

In simulation the *ICI* adaptive window size algorithms demonstrate an improved performance as compared with their nonadaptive window size counterparts.

References

1. W.S. Cleveland and C. Loader, "Smoothing by local regression: principles and methods," in *Statistical theory and computational aspects of smoothing* (Ed.. W. Hardel and M. Schimek), pp. 10–49, Physica-Verlag, 1996.
2. K. Egiazarian, V. Katkovnik, H. Oktem, and J. Astola, "Transform-based denoising with parameters adaptive to unknown smoothness of the signal," in Ed. Creutzburg and Egiazarian: *Spectral Techniques and Logic Design for Future Digital Systems,* Proceedings of International Workshop SPECLOG'2000, Tampere, TTKK, Monistamo, Finland.
3. J. Fan and I. Gijbels, *Local polynomial modelling and its application*, London: Chapman and Hall, 1996.
4. A. Goldenshluger and A. Nemirovski, "Adaptive de-noising of signals satisfying differential inequalities," *IEEE Trans. Inf. Theory,* Vol. 43, No. 3, pp. 873–889, 1997.
5. W. Hardle, *Applied nonparametric regression*. Cambridge, University Press, Cambridge, 1990.
6. T. Hastie and C. Loader, "Local regression: automatic kernel carpentry" (with discussion), *Statistical Science,* Vol. 8, No. 2, pp. 120–143, 1993.
7. C.M. Hurvich and J.S. Simonoff, "Smoothing parameter selection in nonparametric regression using an improved AIC criterion," *Journal of the Royal Statistical Society,* Ser. B, Vol. 60, pp. 271–293, 1998.
8. V. Katkovnik, "A new method for varying adaptive bandwidth selection," IEEE *Trans. on Signal Processing,* Vol. 47, No. 9, pp. 2567–2571, 1999.
9. V. Katkovnik, H. Oktem, and K. Egiazarian, "Filtering heavy noised images using *ICI* rule for adaptive varying bandwidth selection", *ISCAS'99,* Orlando, Florida, May 30 – June 2, 1999.
10. V. Katkovnik, A. Gershman, and L.J. Stankovic, "Sensor array signal tracking using a data-driven window approach", *Signal Processing,* Vol. 80, No. 12, pp. 1507–2515, 2000.
11. V. Katkovnik, K. Egiazarian, and J. Astola, "Median filter with varying bandwidth adaptive to unknown smoothness of the signal," *International Conference on Circuits and Systems (ISCAS'2000),* May 28–31, 2000, Geneva, Switzerland, Proceedings of ISCAS'2000, Vol., pp. 519–522.
12. V. Katkovnik, and LJ. Stankovic, "Periodogram with varying and data?driven window length," *Signal Processing,* Vol. 67, No. 3, pp. 345–358, 1998.
13. V. Katkovnik, *Nonparametric identification and smoothing of data (Local approximation methods).* Nauka, Moscow, 1985 (in Russian).
14. V. Katkovnik, K. Egiazarian, and I. Shmulevich, "Adaptive varying window size filtering based on the intersection of confidence intervals rule," 2001 *IEEE - EURASIP Workshop on Nonlinear Signal and Image Processing,* June 3–6, 2001, Baltimore, Maryland, USA.
15. V. Katkovnik, K. Egiazarian, and J. Astola, "Local transform-based image de-noising with adaptive window size selection," *EOS/SPIE Symposium, Image and Signal Processing for Remote Sensing,* September 25–29, 2000, Barcelona, Spain.
16. C. Loader, *Local regression and likelihood,* Series Statistics and Computing, Springer, N.Y., 1999, pp. 304.
17. S. Mallat, *A Wavelet Tour of Signal Processing,* Academic Press, 1999.
18. J.A. McDonald and Owen A.B., "Smoothing with split linear fits," *Technometrics,* Vol. 28, No. 3, pp. 195–208, 1986.
19. D. Ruppert, "Empirical-bias bandwidths for local polynomial nonparametric regression and density estimation," *JASA*, Vol. 92, No. 439, pp. 1049–1062, 1997.
20. W.R. Schucany, "Adaptive bandwidth choice for kernel regression," *JASA,* Vol. 90, No. 430, pp. 535–540, 1995.

Vladimir Katkovnik received the M.Sc.(1960), Ph.D.(1964) and Doctor of Sc.(1974) in Technical Cybernetics from Leningrad Polytechnic Institute. From 1961 to 1991 he held positions of Assistant, Associate and Professor at the Department of Mechanics and Control Processes of this Institute. From 1991 until 1999 he was a Professor of Statistics Department of the University of South Africa, Pretoria. During 2000 he was a Visiting Professor at Signal Processing Laboratory, Tampere University of Technology, Tampere, Finland.

Currently he is with Mechatronics Department of Kwangju Institute of Science and Technology, South Korea.

His research interests include linear and nonlinear filtering, nonparametric and robust estimation, image processing, nonstationary systems, time-frequency analysis, stochastic optimization and adaptive stochastic control.

He published five books and more than 150 papers.

Karen Egiazarian was born in Yerevan, Armenia, in 1959. He received the M.Sc. degree in mathematics from Yerevan State

University in 1981, and the Ph.D. Degree in physics and mathematics from Moscow M.V. Lomonosov State University in 1986.

In 1994 he was awarded the degree of Doctor of Technology by Tampere University of Technology, Finland. He has been a Senior Researcher at the Department of Digital Signal Processing of the Institute of Information Problems and Automation, National Academy of Sciences of Armenia. He is currently a Full Professor in the Institute of Signal Processing at Tampere University of Technology. His research interests are in the areas of applied mathematics, digital logic, signal and image processing. He has published more than 200 articles in these areas, and is coauthor (with S. Agaian and J. Astola) of the book "Binary Polynomial Transforms and Nonlinear Digital Filters", published by Marcel Dekker, Inc. in 1995, and three book chapters.

Jaakko Astola was born in Helsinki, Finland, in 1949. He received the B.Sc., M.Sc., Licenciate, and Ph.D. degrees in Mathematics from Turku University, Turku, Finland, in 1972, 1973, 1975, and 1978, respectively.

From 1976 to 1977, he was a Research Assistant at the Research Institute for Mathematical Sciences of Kyoto University, Kyoto, Japan. Between 1979 and 1987, he was with the Department of Information Technology, Lappeenraanta University of Technology, Lappeenraanta, Finland, holding various teaching positions in mathematics, applied mathematics, and computer science. From 1988 to 1993, he was an Associate Professor in applied mathematics at Tampere University, Tampere, Finland. Currently, he is a Professor of digital signal processing at Tampere University of Technology, the head of Signal Processing Laboratory and the Director of the Tampere International Center in Signal Processing (TICSP). His research interests include signal processing, coding theory, and statistics.