

Point-Wise Adaptive Wavelet Transform for Signal Denoising

Mladen TOMIC^{1*}, Damir SERSIC²

¹*University of Rijeka, Faculty of Engineering
Vukovarska 58, 51000 Rijeka, HR-Croatia*

²*University of Zagreb, Faculty of Electrical Engineering and Computing
Unska 3, 10000 Zagreb, HR-Croatia*

e-mail: mladen.tomic@riteh.hr, damir.sersic@fer.hr

Received: March 2012; accepted: October 2012

Abstract. Underperformance in higher frequency signal regions denoising is a common problem for many denoising methods. Wavelet transforms are, generally, less prone to the problem than the pure spatial or frequency domain transforms, but there is still much room for improvements. In this paper, we propose a point-wise adaptive wavelet transform for signal denoising applications. It is very efficient in denoising higher frequency regions, without compromising the performance on smooth, lower frequency, regions. The transform uses statistical method of intersection of confidence intervals rule to adapt to local signal properties. Its performance was extensively tested on various signal classes. The results proved validity of theoretical assumptions and showed significant performance improvements when compared to other denoising methods.

Key words: point-wise adaptation, wavelet selection, adaptive lifting scheme, adaptive wavelet transform, ICI rule, signal denoising, edge-preserving denoising.

1. Introduction

Most of the real-world signals are corrupted by noise. Hence, it is necessary to employ efficient signal denoising algorithms to improve signal quality. Wavelet transforms were proved to be a valuable tool in signal denoising. Popularity of wavelet transforms was further increased after the ladder structure (Sweldens, 1996) for transform calculation was proposed. The structure is called the lifting scheme. The lifting scheme allows for more efficient transform calculation, makes the perfect reconstruction property guaranteed, performs all calculations in the spatial domain and makes it easy to introduce adaptivity to a transform.

1.1. Adaptive Wavelet Transforms

Most denoising algorithms perform well on smooth signals or smooth signal regions. The greatest difference in the algorithm efficiency can be noted for signals containing higher

* Corresponding author.

frequency features, such as discontinuities, spikes, bumps, or any other sudden changes in local signal properties. Compactly supported wavelets stand out as an excellent performer for such signal classes, but there is still much room for improvements.

The scale-adaptive (ScAT) and the space-adaptive (SpAT) lifting schemes were proposed in Claypoole *et al.* (2003). ScAT performs adaptation on a scale-by-scale basis, while SpAT adapts on a point-by-point basis. From a set of predictors, one which results in the smallest prediction error is chosen. To eliminate the necessity of bookkeeping due to a space-varying predictors, the update first framework was also proposed. Unusual approach of adapting the update filter, based on decision maps was proposed in Piella *et al.* (2002), Piella and Heijmans (2002). A simple adaptation method, in which only two conventional wavelets are used, was presented in Wu *et al.* (2004), in which choice of the wavelet depends on a straightforward step-edge detection algorithm. Interesting approach was taken in Chan and Zhou (1999). Instead of modifying the transform, signal itself is modified at discontinuities. After the discontinuity is detected, signal values from one side of the discontinuity are used to extrapolate its values to the other side of the discontinuity, resulting in a smooth signal. After denoising, a reverse filters are used to restore discontinuities. Finally, similar approach to the one taken in this paper was explored in Sersic (2000). The basic idea was to use filter banks with more vanishing moments for smoother parts of a signal, while for the transients and singularities, filter banks with less vanishing moments were to be used.

In a separate research field, a successful denoising methods built upon local polynomial approximation (LPA) and intersection of confidence intervals (ICI) rule were proposed. The ICI rule statistical method was used as a tool for adaptive support selection. First the 1-D case was examined (Katkovnik, 1999) and afterwards the ICI was used for very efficient image denoising algorithms (Foi *et al.*, 2007; Katkovnik, 2003). In attempt to further improve the efficiency of the ICI method, relative ICI (RICI) method was proposed in Lerga *et al.* (2011).

In this paper, we propose an adaptive denoising method based on a lifting scheme, which efficiently utilizes the ICI to adapt the transform. By incorporating the ICI into the lifting scheme, it becomes possible to very efficiently adapt to local signal properties, on a point-by-point basis, as briefly presented in Tomic and Sersic (2012). Instead of focusing on certain features detection, such as a discontinuities, the ICI allows us to detect neighborhood of each point, which share similar statistical properties. This way, on each scale and for each point, it becomes possible to select an appropriate wavelet basis, such that the lifting filters do not span across two or more regions of different local properties. The approach yields in significant reduction of visible denoising artifacts and general improvement in denoising performance. The proposed method can be used as a standalone 1-D signal denoising method, as a part of more complex algorithms, such as medical image denoising algorithm (Tomic *et al.*, 2012a) or as a preprocessing step in many signal analysis applications, such as Su (2011).

2. Rationale

A large number of wavelet basis functions exists. Some important properties of wavelets, from the signal denoising point of view, are the support size, smoothness and symmetry.

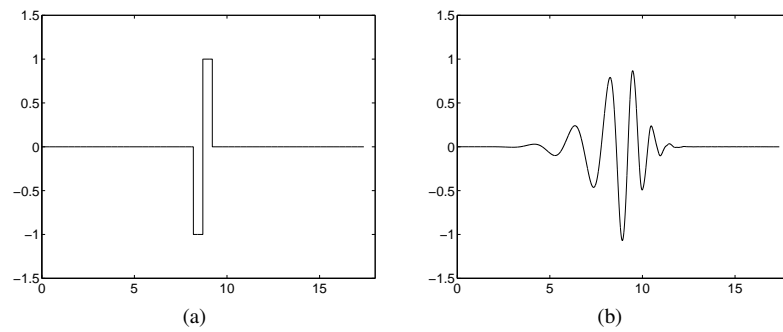


Fig. 1. Wavelet functions: (a) Haar, (b) Daubechies 9.

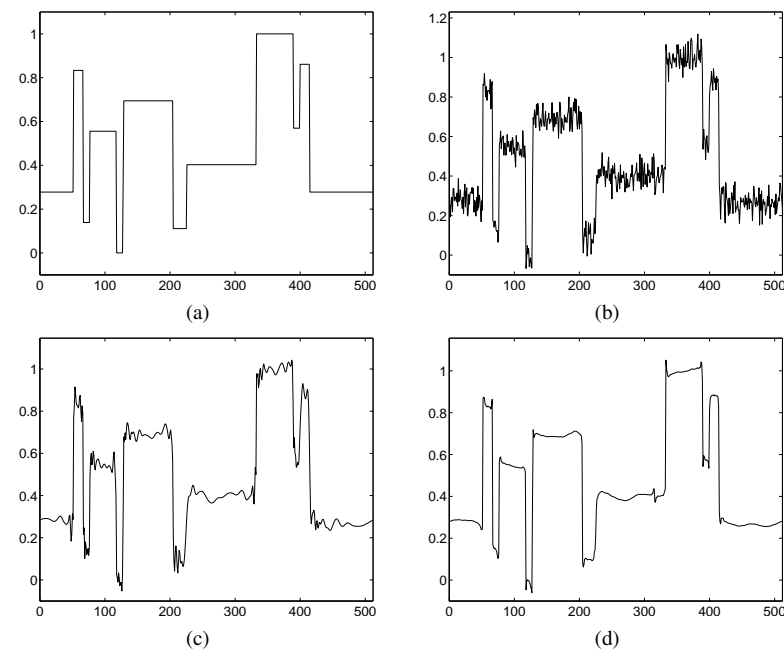


Fig. 2. Blocks signal denoising: (a) original signal, (b) noisy signal, (c) signal denoised using Db9 and (d) signal denoised using Haar wavelet.

For instance, two different wavelets are shown in Fig. 1 – Haar and Daubechies 9 (*Db9*). Haar wavelet features short support and a step edge, while the *Db9* has much longer support, it is oscillatory and very smooth. Since reconstructed signal can be viewed as a superposition of coarse signal approximation and scaled and translated wavelet functions, it is easy to predict the effect of using either of the shown wavelets to denoise some typical signals. Scaled and translated Haar wavelet very efficiently reconstructs sharp edges but has difficulties with smooth signal regions. Smoothness and longer support of the *Db9* wavelet represent the opposite case. It efficiently reconstructs smooth signal regions, but is not able to do the same with higher frequencies in a signal, eventually leading to a pseudo-Gibbs oscillations at sharp edges. This is illustrated in Fig. 2.

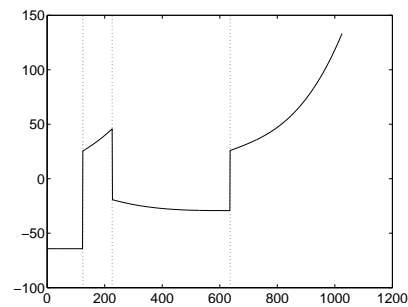


Fig. 3. Sample signal with variable local properties.

The example points out general principle for wavelet basis selection. If a given signal is characterized by smooth and slow transitions, i.e. low frequencies, longer and smoother wavelet basis should be chosen. If the signal is characterized by high frequencies, shorter wavelet should be chosen. Even for a class of stationary signals, there is no simple solution to decide the exact optimal wavelet shape or support length. Non-stationary signals represent a more difficult scenario, as varying signal properties means it would be best to use several different wavelet basis for different signal regions.

2.1. Adaptive Lifting Scheme Basic Concept

Let us examine the signal in Fig. 3. It features four smooth segments, connected by three sharp edges. Our adaptive algorithm should ensure that the longest wavelet which does not span the edge in a signal is used for each signal point. For instance, at the edge samples, *Haar* wavelet should be used as it allows for excellent edge reconstruction. As we move away from the edge, longer wavelets should be used to allow for fine reconstruction of smooth signal regions. If done right, the algorithm will combine good properties of different wavelet basis into a single transform. Denoising scheme based on the transform should perform well on both smooth regions and edges, contained in the same signal.

2.2. Definition and Detection of an Edge

Strategy of changing the wavelet basis as we approach the edge, or move away from it, brings about another obstacle – edge detection. Edge detection is a problem per se. Before trying to implement any of the detection algorithms, it is necessary to define what does the *edge* exactly mean to our application.

It could be defined as a step edge, a spike or even any high frequency change (drawing another question: What is a high frequency?). In this particular application we assume that the signal can be modeled as a piece-wise polynomial function. We find the usable definition of an edge to be: a point in a signal which divides two polynomial segments. It is a rather broad definition, which includes much more than what common edge detectors are trying to detect. To detect the existence of edges conforming to the given definition, we use the statistical method presented in Katkovnik (1999) – Intersection of Confidence Intervals (ICI) rule. As explained in more details later, the ICI rule helps us determine the largest neighborhood about each signal point, which shares the same local properties.

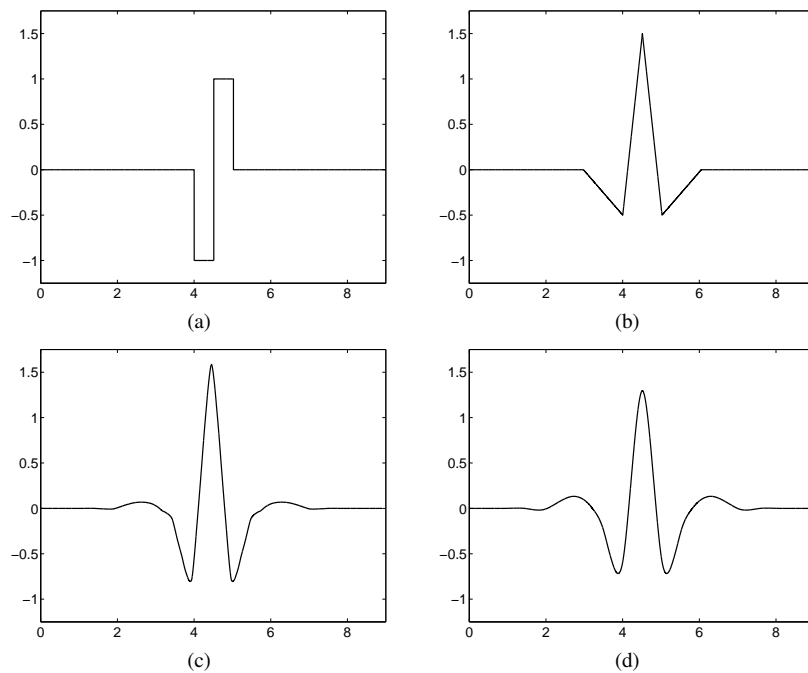


Fig. 4. Biorthogonal wavelet set for adaptive algorithm: (a) *Haar* (*Bior1.1*), (b) *Bior2.2*, (c) *Bior4.4* and (d) *Bior6.6*.

2.3. Wavelet Basis Set

It was shown in Daubechies and Sweldens (1998) that any wavelet transform with finite filters can be decomposed into a finite sequence of simple lifting steps. Having many lifting steps leads to an unnecessary algorithmic and computational complexity, while adding little benefits. Out of the many existing wavelet families, we propose the biorthogonal wavelets, which are “native” to the lifting scheme, to be used. Only two lifting steps are required to construct such wavelets: a single predict and a single update step.

Figure 4 shows 4 wavelets used in the adaptive scheme. The wavelets are: *Haar*, *Bior2.2*, *Bior4.4* and *Bior6.6*. Should the adaptive algorithm perform well, the *Haar* wavelet will be used about sharp edges in a signal, while the *Bior6.6* will be used for the smoothest signal regions.

Based on the results of the ICI rule, a set of lifting filters, constructing one of the four wavelets, is chosen and used at each signal point. The adaptation is performed independently for each point and on each scale.

3. Intersection of Confidence Intervals – ICI

The ICI is a statistical method which utilizes non-parametric local polynomial approximation (LPA) and confidence intervals (CI) theory to estimate the real value of a sample in a noisy signal. We briefly present part of the method theory which is relevant and mod-

ified for the adaptive lifting scheme, while the complete presentation and analysis of the method can be found in Katkovnik (2003).

Let the signal:

$$y(x), \quad y \in \mathbb{R}^1, \quad x \in \mathbb{R}^1 \quad (1)$$

be a piece-wise polynomial function, sampled on a regular or irregular grid:

$$x = x(n),$$

where index n corresponds to the n -th signal sample. Its noisy observations $z(x)$ are given by:

$$z(x) = y(x) + \epsilon(x), \quad (2)$$

where $\epsilon(x) \sim \mathcal{N}(0, \sigma_\epsilon^2)$ is considered to be independent Gaussian white noise.

The aim of the method is to estimate real values of the unknown deterministic $y(x)$, from its noisy observations $z(x)$. Each $y(x)$ is considered to be lying on a polynomial segment. The segment can be approximated by a polynomial of order $m = 0, \dots, M$. We are only interested in the polynomial value at a given point, so, instead of finding the polynomial parameters, the non-parametric LPA route is taken.

3.1. Local Polynomial Approximation – LPA

Let the point x be a center point of the LPA, i.e., the point at which the function value $y(x)$ is to be estimated using the LPA kernels. Estimate for the point x_s in the neighborhood of x can be looked for as follows:

$$\begin{aligned} y(x - x_s) &= C^T u(x - x_s), \\ u(x) &= (u_1(x), u_2(x), \dots, u_M(x))^T, \\ C &= (C_1, C_2, \dots, C_M)^T, \end{aligned} \quad (3)$$

where $u(x) \in R^M$ is a vector of linearly independent 1D polynomials, of the power from 0 to M , while $C \in R^M$ is a vector of parameters for this model.

In order to find the coefficients C from (3), the following criterion function is applied:

$$J_h(x) = \sum_s w_h(x - x_s) (z(x_s) - y(x - x_s))^2, \quad (4)$$

where the window $w_h(x)$:

$$w_h(x) = \frac{w(x/h)}{h^2}, \quad \int w_h(x) dx = 1, \quad (5)$$

formalizes localization of fitting with respect to the center x , while the scale parameter $h > 0$ determines the window size.

Estimates of parameters C are obtained by minimizing the criterion function (4) with respect to the C :

$$\begin{aligned}\hat{C}(x, h) &= u_h^{-1} \sum_s w_x(x - x_s) u(x - x_s) z(x_s), \\ u_h &= \sum_s w_h(x - x_s) u(x - x_s) u^T(x - x_s), \\ \hat{C}(x, h) &= (\hat{C}_1(x, h), \dots, \hat{C}_M(x, h))^T.\end{aligned}\quad (6)$$

Having obtained coefficients \hat{C} , estimate of the function value at point x_s follows, from (3) and (6), as:

$$\begin{aligned}\hat{y}_h(x - x_s) &= \hat{C}^T u(x - x_s) = u^T(x - x_s) \hat{C} \\ &= u^T(x - x_s) u_h^{-1} \sum_s (w_h(x - x_s) u(x - x_s) z(x_s)).\end{aligned}\quad (7)$$

Let us assume that the polynomials $u(x - x_s)$ are orthonormalized with respect to the window $w_h(x - x_s)$. We shall denote the orthonormalized set of polynomials as $\phi(x - x_s)$. The expression (7) can now be written as:

$$\hat{y}_h(x - x_s) = \phi^T(x - x_s) \phi_h^{-1} \sum_s (w_h(x - x_s) \phi(x - x_s) z(x_s)).\quad (8)$$

Considering that $\phi_h = I_{M \times M}$, and substituting $\phi^T(d_s) = \phi^T(x - x_s)$, expression (8) becomes:

$$\hat{y}_h(x - x_s) = \sum_s (w_h(x - x_s) \phi^T(d_s) \phi(x - x_s) z(x_s)),$$

which can be further simplified to:

$$\begin{aligned}\hat{y}_h(x - x_s) &= \sum_s g_h(x - x_s) z(x_s), \\ g_h(x - x_s) &= w_h(x - x_s) \phi^T(d_s) \phi(x - x_s),\end{aligned}\quad (9)$$

showing that the estimate can be obtained by filtering the noisy signal with the FIR filter, given by its impulse response $g_h(x - x_s)$.

3.2. Idea of the ICI

Let us consider the LPA kernels using estimation order m and window of size h . In case the center point of LPA lies on a polynomial segment whose order is $\leq m$, the kernels provide perfect estimate of the function real value, for any window size $h > m$. In reality, the original signal $y(x)$ is composed of many different polynomials and also corrupted by

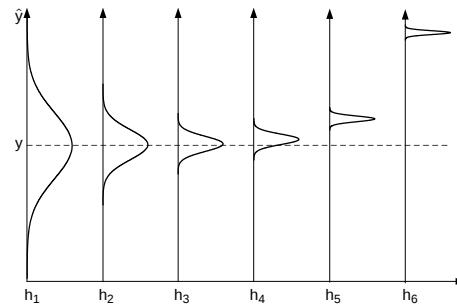


Fig. 5. Probability distribution functions of $\hat{y}(x)$, for growing window sizes h_i .

noise. Perfect reconstruction is highly unlikely, but widening of the estimation window can improve $\hat{y}(x)$ accuracy and lower its variance.

The problem with growing estimation window arises when the enlarged window includes points which lie on a different polynomial segment. LPA kernel estimators become biased and variance alone cannot be looked upon as a measure of accuracy. Should the window be further extended, estimation bias could cause significant worsening of the $\hat{y}(x)$, regardless of the low estimate variance. The effect is depicted in Fig. 5.

It is evident that an unbiased estimator can result in an unacceptably high estimate variance, while at the same time lowering of the variance, by extending the estimation window, can result in an unacceptable estimation bias. The ICI rule considers both of the variables and, as shown in Stankovic (2004), provides optimal bias-to-variance trade-off.

3.3. Determining Support Size – ICI Rule

Let us examine confidence intervals (CI) for the $\hat{y}(x)$. Their upper (U) and lower (L) limits are given by:

$$\begin{aligned} U &= \hat{y}(x) + z_c \cdot \sigma_{\hat{y}}, \\ L &= \hat{y}(x) - z_c \cdot \sigma_{\hat{y}}, \end{aligned} \quad (10)$$

where the critical level $z_c = \chi_{1-\alpha/2}$ is the $(1 - \alpha/2)$ -th quantile of the standard Gaussian distribution and defines the probability that the true value $y(x)$ is contained within the CI limits.

To make the expression (10) more accurate, the bias has to be considered. The upper bound of the estimation bias (Stankovic, 2004) is given by:

$$\begin{aligned} \bar{\omega}(x, h) &< \gamma \cdot \sigma_{\hat{y}}, \\ \gamma &= (\sqrt{m+1})^{-1}. \end{aligned} \quad (11)$$

Then, from (10) and (11), the expressions for CI limits are:

$$\begin{aligned} U &= \hat{y}(x) + \Gamma \cdot \sigma_{\hat{y}}, \\ L &= \hat{y}(x) - \Gamma \cdot \sigma_{\hat{y}}, \end{aligned} \quad (12)$$

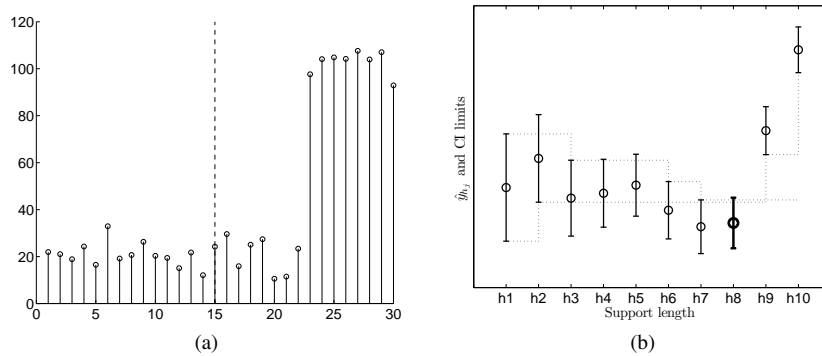


Fig. 6. CIs for $\hat{y}(15)$. Filter support is growing to the right: (a) sample noisy signal and (b) CI limits.

where $\Gamma = \gamma + \chi_{1-\alpha/2}$ is a free parameter, which defines the method sensitivity. Too large or too small parameter Γ value results in, respectively, oversmoothing or undersmoothing the signal.

Let us calculate CIs for estimators of growing estimation windows, i.e., growing supports: $\mathbb{H} = \{H_i, i = 1, \dots, N\}$, with length h_i of each succeeding support being larger than the previous one: $\{h_i \mid h_i < h_{i+1}\}$. Maximum lower and minimum upper CI limits are recorded for each H_i (Katkovnik *et al.*, 2002):

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

The ICI rule states that the estimator producing the most accurate estimate $\hat{y}(x)$ should be the one with the largest support H_i , for which the condition:

$$\bar{L}_i \leq \underline{U}_i \tag{13}$$

is still satisfied. The respective support shall be denoted as H^+ . It is to be expected that any support larger than H^+ would include signal points which belong to a region with different local properties. This is illustrated in Fig. 6. It can be seen in Fig. 6(b) that the $y(15)$ should be estimated using H_8 , as it is the largest filter support for which the condition (13) still holds.

4. Edge Preserving Lifting Scheme – ICI-EPL

Let us consider the lifting scheme realization of the wavelet transform (Fig. 7). In the predict step, filter P is used to predict values of the odd-indexed signal sample, based on a certain number of even-indexed samples, and vice versa. Because, for conventional wavelets, filter P is same for every signal sample, we get a denoising performance trade-off – longer filter P results in better performance in smooth signal regions, while shorter P

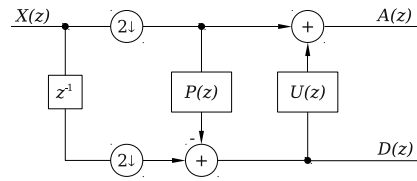


Fig. 7. Single level lifting scheme decomposition.

results in better performance about edges in a signal. In order to avoid the trade-off, the ICI rule can be used to determine the filter P support for each signal sample, independently.

4.1. Lifting Filters

As explained in Section 3.3, the ICI rule considers LPA kernels (9), while lifting filters used in the lifting scheme are Neville filters (Kovacevic and Sweldens, 2000). There are two options to consider – either Neville filters will be used for the ICI rule, or the LPA type kernels will be used in the lifting scheme. Let us briefly examine both options.

4.1.1. Neville Filters and the ICI Rule

Essentially, Neville filters of order N are also LPA kernels. They are interpolating filters which are able to perfectly reconstruct any polynomial whose order is $m < N$. The same is true for the LPA kernels. Difference between the ICI utilization of LPA kernels and Neville filters is in the relationship of order of estimation and support size.

LPA kernels used for the ICI rule all share the same estimation order, set to a value of $m = \text{const}$, regardless of the kernel support size. On the other hand, as the support of Neville filters grow, so does their order of estimation, which is always equal to: $m = h - 1$. If used for the ICI rule, both filter types must have growing support, but in the LPA kernel case, estimation order is fixed, while for the Neville filters it grows with the support. Important consequence is relationship of estimate variance to support size. Both estimator types are FIR filters and the respective estimate variances are given by:

$$\begin{aligned} \hat{y}(x) &= \sum_k g_h(k) y(x - k), \\ \text{Var}(\hat{y}) &= \sigma_{\hat{y}}^2 = \sum_k g_h^2(k) \cdot \sigma_{y(x-k)}^2. \end{aligned} \quad (14)$$

Considering that $\sigma_{y(x-k)} = \sigma_{\epsilon}$, the variance follows as:

$$\text{Var}(\hat{y}) = \sigma_{\hat{y}}^2 = \sigma_{\epsilon}^2 \cdot \sum_k g_h^2(k). \quad (15)$$

The variance is directly proportional to the sum of squared filter/kernel coefficients, and, as per (10), the confidence interval width is directly proportional to the variance. In the LPA kernel case the variance and the CI width are getting lower as the support grows,

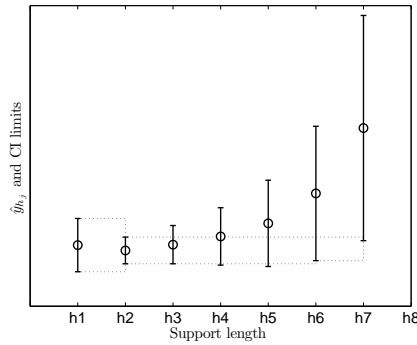


Fig. 8. Effect of growing variance of Neville filters on the ICI rule efficiency.

while for the Neville filters, they are getting higher. The ICI rule efficiency is based on the shrinking CI widths, as shown in Fig. 6. The effect of expanding CI widths can be observed in Fig. 8. The figure shows plot of CI bars for a sample signal, when using Neville filters of order $N \in 1, 2, 3, 4, 5, 6, 7$. It can be seen that the expanding CIs are detrimental to the efficiency of the ICI rule, as it slows or even disables the breaching of the condition (13).

4.1.2. LPA Kernels and the Lifting Scheme

LPA kernels as proposed in Katkovnik (1999) cannot be directly used in the lifting scheme. It was assumed that the kernels use every support point for the approximation. Also, in (9), the center point of the LPA was assumed to always be equal to x_s . It leads to $d_s = 0$ and to a modified expression:

$$g_h(x - x_s) = w_h(x - x_s)\phi^T(0)\phi(x - x_s). \tag{16}$$

Because of the assumptions, the LPA kernels are not suitable for inclusion into the lifting scheme, as it presumes a polyphase decomposition. Minor modifications to the original approach are necessary to generate kernels which use samples from only one signal phase to predict values of samples from the other signal phase. Let us introduce a rectangular asymmetrical window function:

$$w_h(x - x_s) = 1/h,$$

and assume all points of the support H to belong to the same signal phase. Samples from the other phase now lie at the half-integer grid points. To define the point whose value is to be estimated, the d_s in (9) shall be set to a proper half-integer value. For instance, given a support length $h = 4$, and setting the $d_s = 1.5$ results in a symmetrical kernel, estimating the point in the center of the support. Such kernels can now be upsampled and used in the lifting scheme as predictors P .

Problem with the approach is that the primary purpose of filter P is wavelet construction, and not only achieving a good prediction, so we must examine the wavelets which get constructed using the LPA kernels. Figure 9 shows wavelets in case LPA kernels with

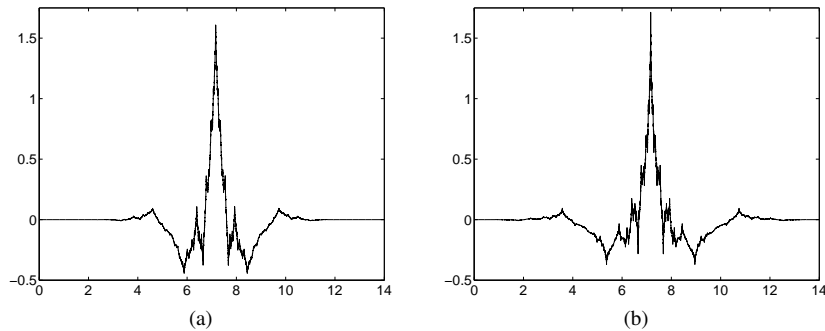


Fig. 9. Synthesis wavelets from LPA kernels with estimation order $m = 2$. Support length: (a) $h = 6$ and (b) $h = 8$.

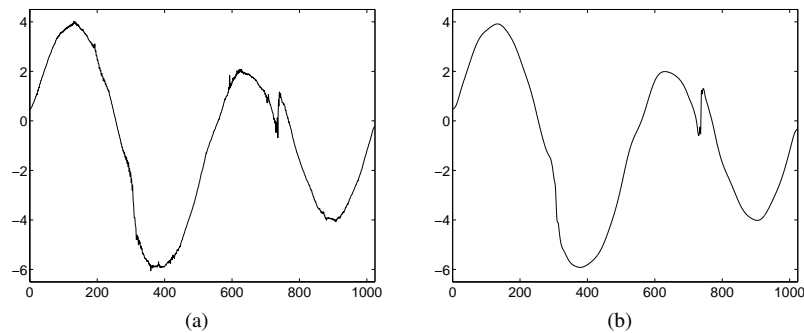


Fig. 10. *HeaviSine* signal denoised using wavelet transforms based on: (a) LPA kernel, $m = 2$, $h = 8$ and (b) Neville filter $h = 8$.

estimation order $m = 2$ are used. Plots are provided for support lengths $h \in \{6, 8\}$. If we compare the resulting wavelets with the ones from Neville filters, shown in Fig. 4, we can see some general similarity in shape but there is a significant difference in regularity. Wavelets constructed using Neville filters are very smooth, while the wavelets from LPA kernels contain numerous singularities.

Given that the reconstructed signal is a sum of the coarse signal approximation and translated and scaled wavelet functions, negative influence of such wavelets on general denoising performance is easily anticipated. It would be difficult to reconstruct smooth signal regions without denoising artifacts in form of, mostly, small fluctuations about the real signal value, being clearly visible. The effect is shown in Fig. 10. The example compares two denoised *HeaviSine* signals. One is denoised using the wavelets resulting from Neville filters with support length $h = 8$, while the other is denoised using the wavelets resulting from LPA kernels with $m = 2$ and the same support length, $h = 8$.

4.2. Combining the Advantages

We showed that the Neville filters work excellent for wavelet construction but are not suitable to be used for the ICI rule. On the other hand, LPA kernels can be used with the ICI rule to achieve efficient support size determination, but wavelets they construct are

inferior to the ones constructed by Neville filters. To explore the advantages of both filter types, we propose the signal decomposition in the lifting scheme to be broken into two separate steps.

In the first step, on each scale and for each point, the ICI rule with the symmetrical LPA estimators of order m , with set of growing supports:

$$\mathbb{H}_r = \{H_{m,i}, i = 1, \dots, N\}$$

of length:

$$\{h_{m,i} \mid h_{m,i} < h_{m,i+1}\}$$

are used to find the largest admissible support (denoted by '+') H_m^+ of length h_m^+ . Based on the H_m^+ even-indexed points (denoted as $H_{m,e}^+$), the estimators predict value of one of the H_m^+ odd-indexed points (denoted as $H_{m,o}^+$), and vice versa. Since symmetrical estimators are assumed, the supports $H_{m,i}$ also grow symmetrically. Ideally, samples which comprise the support H_m^+ would not belong to two (or more) signal regions with different local properties. Asymmetrical estimators are not considered as they have negative impact on overall denoising performance.

In the second step, actual decomposition is performed by substituting the chosen LPA kernels with Neville filters of the same support H_m^+ . The substitution can be justified by the fact that the resulting smooth wavelets will have at least the same number of vanishing moments as if the LPA kernel estimators were used. Possible additional vanishing moments would only provide smoother reconstruction but would have no negative impact on edge preservation.

4.2.1. Choosing the Estimation Order

Basically, any estimation order m can be chosen for the LPA kernels, although differences in actual denoising performance are not very significant. Another factor to consider when choosing the estimation order is a potential for support growth. Because employing longer wavelets do not bring further performance improvements, we have decided the longest wavelet to be the *Bior6.6*. In effect, the longest filter support that can be used for the ICI rule is of length $h = 6$. Too large estimation order m would, thus, limit the potential for support growth. For instance, should we chose the order $m = 3$, minimum support length would have been $h = m + 1 = 4$. Since the support is grown symmetrically, only two filters would have been considered by the ICI algorithm, which is not enough to explore its benefits or ensure reasonable reliability.

As well as limiting support growth potential, larger estimation order values also entirely exclude shorter wavelets. It means that longer wavelets would be used even at step edges. This is in harsh contradiction with the basic idea of adaptation and can severely lower transform performance about edges in a signal.

Simulations showed that the estimation order $m = 1$ proved to be enough for optimal denoising performance so it is the order we propose to be used in the ICI-EPL.

4.2.2. Support Verification

Since we opted to set the kernel estimation order to $m = 1$, the resulting symmetrical estimators are simple averaging filters:

$$P_N(z) = \frac{1}{N} \sum_{k=0}^{N-1} z^{-k}.$$

As a result, the more the support grows, the lesser is the impact of added samples ($H_{m,i+1} \setminus H_{m,i}$) on the prediction result. It means that some edges might still show up in the final support H_m^+ . Small denoising performance improvements can be achieved by further examining the H_m^+ .

In order to verify the support and increase probability that it does not contain any sudden changes in signal statistics, two asymmetrical LPA kernel estimator sets are introduced. Each set is independently used in the ICI algorithm as a set of filters with growing supports. They predict values of the, respectively, second and next to the last points of H_m^+ . Their supports are growing from boundaries of the H_m^+ to the right and to the left, respectively. The ICI algorithm results in supports H_m^{L+} and H_m^{R+} , whose lengths (h_m^{L+} and h_m^{R+}) are $\leq h_m^+$. The asymmetrical estimators ensure that samples added in each step of the ICI algorithm carry higher weight, which increases the probability of recognition of changes in local signal properties.

In case that:

$$\min[h_m^{L+}, h_m^{R+}] < h_m^+,$$

it is possible that the ICI rule did not indicate change in signal properties fast enough and that the H_m^+ contains signal samples from different segments. Our strategy is to contract the H_m^+ by one filter tap from both sides. The verification step is then repeated until:

$$h_m^{L+} = h_m^{R+} = h_m^+.$$

5. Results and Discussion

To investigate performance of the proposed denoising scheme, we use the usual test signals: *Blocks*, *Bumps*, *Doppler*, *HeaviSine*, *Piece-Polynomial* and *Piece-Regular*, and the *Blocks-HeaviSine* composite signal. Additive white Gaussian noise was added to each of the signals. Standard deviations of $\sigma_\epsilon \in \{5\%, 10\%, 15\%\}$ of total signal magnitude were considered.

We compared the ICI-EPL to the original ICI method, as proposed in Katkovnik (1999), and to 5 widely used conventional wavelet transforms: *Haar*, *Bior2.2*, *Bior4.4*, *Db3* and *Db9*. To eliminate influence of threshold selection on the performance comparison, we used oracle thresholding, i.e. the optimal threshold value, for each of the wavelet transforms. The ICI-EPL parameter Γ value was chosen the same way.

Table 1
RMSE values of denoised signals for additive Gaussian noise with $\sigma_\epsilon = 5\%$ of total signal magnitude.
RMSE multiplied by 10^2 .

	Wavelets	ICI	ICI-EPL
Blocks	1.22–2.70	2.20	0.89
Bumps	1.91 –2.53	2.25	1.94
Doppler	1.66 –2.49	2.01	1.69
HeaviSine	1.14–1.41	1.14	1.09
Piece-Polynomial	1.21–2.47	1.73	0.95
Piece-Regular	1.72–2.32	1.80	1.41
Blocks-HeaviSine	1.31–2.15	1.57	1.10

Table 2
RMSE values of denoised signals for additive Gaussian noise with $\sigma_\epsilon = 10\%$ of total signal magnitude.
RMSE multiplied by 10^2 .

	Wavelets	ICI	ICI-EPL
Blocks	2.71–4.85	3.81	2.25
Bumps	3.67 –4.67	4.00	3.73
Doppler	3.15–4.12	3.77	3.09
HeaviSine	1.91 –2.23	2.29	2.08
Piece-Polynomial	2.50–4.27	3.35	2.22
Piece-Regular	3.12–4.27	3.20	2.96
Blocks-HeaviSine	2.40–3.77	2.71	2.25

Table 3
RMSE values of denoised signals for additive Gaussian noise with $\sigma_\epsilon = 15\%$ of total signal magnitude.
RMSE multiplied by 10^2 .

	Wavelets	ICI	ICI-EPL
Blocks	4.09–6.54	5.04	3.67
Bumps	5.21 –6.36	5.64	5.33
Doppler	4.43–5.57	5.24	4.36
HeaviSine	2.58 –2.93	3.18	2.85
Piece-Polynomial	3.71–5.53	4.75	3.63
Piece-Regular	4.50–5.63	4.63	4.47
Blocks-HeaviSine	3.49–4.97	3.68	3.43

5.1. Objective Quality – RMSE

Objective performance measures are given in terms of the root mean square error (RMSE) and presented in Tables 1–3. In case of signals which contain singularities, it can be seen that the ICI-EPL easily, and in some cases by a significant margin, outperforms other considered transforms. In all other cases, the ICI-EPL has a performance comparable to the best achieved denoising result.

5.2. Subjective Quality – Visual Inspection

RMSE is a commonly used objective measure for denoising performance, however, it is also beneficial to visually inspect denoised signals, as there is often a gap between objective and subjective perception of denoised signal quality.

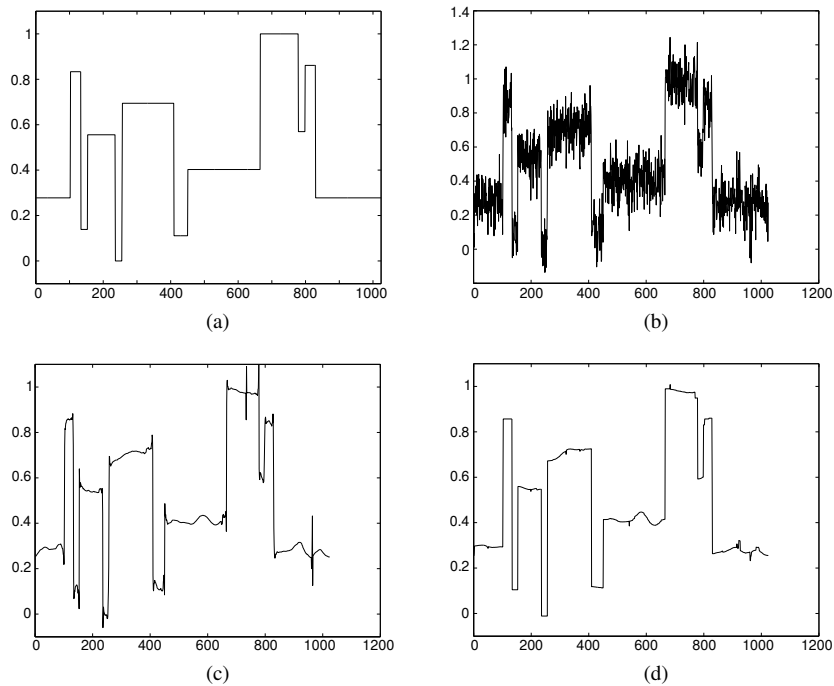


Fig. 11. *Blocks* signal denoising, $\sigma_\epsilon = 10\%$ of total signal magnitude: (a) original signal, (b) noisy signal, (c) signal denoised using optimal wavelet transform (*Haar*) and (d) signal denoised using ICI-EPL.

5.2.1. *Blocks* Signal

Blocks signal, Fig. 11(a), is a piece-wise constant signal, characterized by sharp step edges between constant-value signal regions. Of the fixed wavelets, *Haar* wavelet is optimal transform for denoising, as its shape perfectly fits the signals step edges. Figure 11(b) shows noisy signal ($\sigma_\epsilon = 10\%$ of total signal magnitude), while Figs. 11(c) and (d) show, respectively, signals denoised using the *Haar* wavelet and the ICI-EPL. The ICI-EPL advantage of $\approx 20\%$ in RMSE is also clearly visible in denoised signal plots. The ICI-EPL shows exceptional edge reconstruction and reconstruction of shorter duration constant-value regions, while in case of the *Haar* wavelet, denoised signal exerts visible spikes and roughnesses near edges.

5.2.2. *Bumps* Signal

Bumps signal, Fig. 12(a), is characterized by a series of bumps (or spires). Except near the bump peaks, the signal can be considered to be smooth. Figure 12 shows denoising results for a high noise level case – $\sigma_\epsilon = 15\%$ of total signal magnitude. Of the tested conventional wavelets, the *Bior2.2* performed the best for this type of signal. In terms of the RMSE, the ICI-EPL scores $\approx 2.3\%$ worse result than the *Bior2.2*, but in terms of the visual quality of denoised signal, a different conclusion can be made.

ICI-EPL performance at zero value regions is comparable to the *Bior2.2* performance, although there are several unwanted spikes. Still, as can be seen in Fig. 12(d), the false

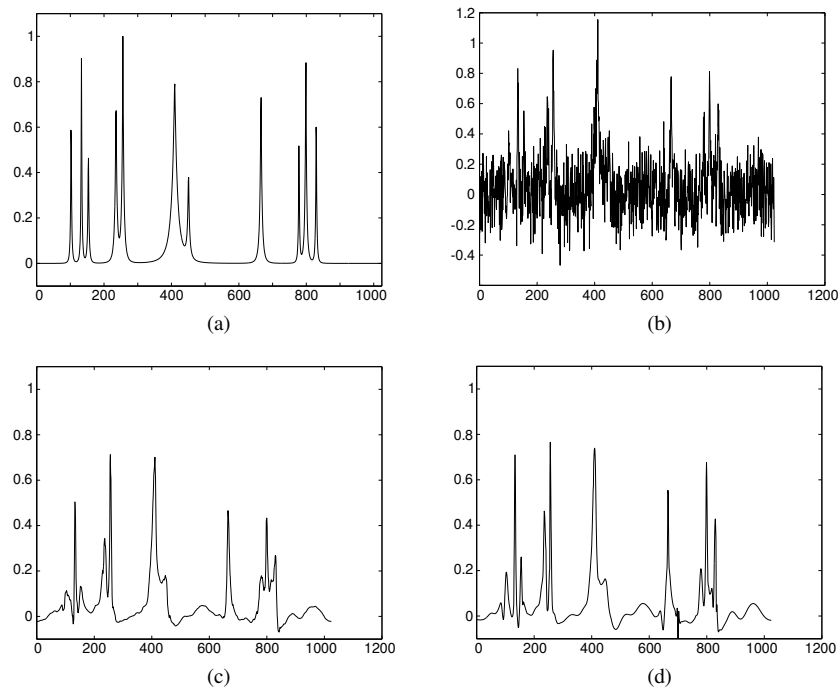


Fig. 12. *Bumps* signal denoising, $\sigma_\epsilon = 15\%$ of total signal magnitude: (a) original signal, (b) noisy signal, (c) signal denoised using optimal wavelet transform (*Bior2.2*) and (d) signal denoised using ICI-EPL.

spikes are easily distinguished from the real bumps in a signal. When we compare the bumps reconstruction, ICI-EPL performance is far superior. The first series of bumps is badly corrupted in the *Bior2.2* case, while the ICI-EPL successfully conserved all three bumps in the series. Similar observation can be made for the fourth series of bumps, where, unlike the *Bior2.2* case, all three bumps can be distinguished in the ICI-EPL case. Also, bump magnitudes in the ICI-EPL denoised signal are much closer to the original signal magnitudes.

Following above considerations and a fact that the bumps in the *Bumps* signal carry most of the useful signal information, we can conclude that the ICI-EPL performed significantly better in the presented high noise scenario, although the RMSE suggests slightly lower performance.

5.2.3. *Blocks-HeaviSine* Composite Signal

The last signal to be inspected is a composite *Blocks-HeaviSine* signal, Fig. 13. By combining the two signals, we can test how well the ICI-EPL is able to adapt to a substantially different local properties of the same signal – step edges of the *Blocks* signal part and mostly smooth shapes of the *HeaviSine* signal part. Figure 13 shows denoising results, for $\sigma_\epsilon = 5\%$ of total signal magnitude. The ICI-EPL achieved $\approx 18\%$ lower RMSE than the second best transform – *Haar* wavelet. The large difference in performance can also

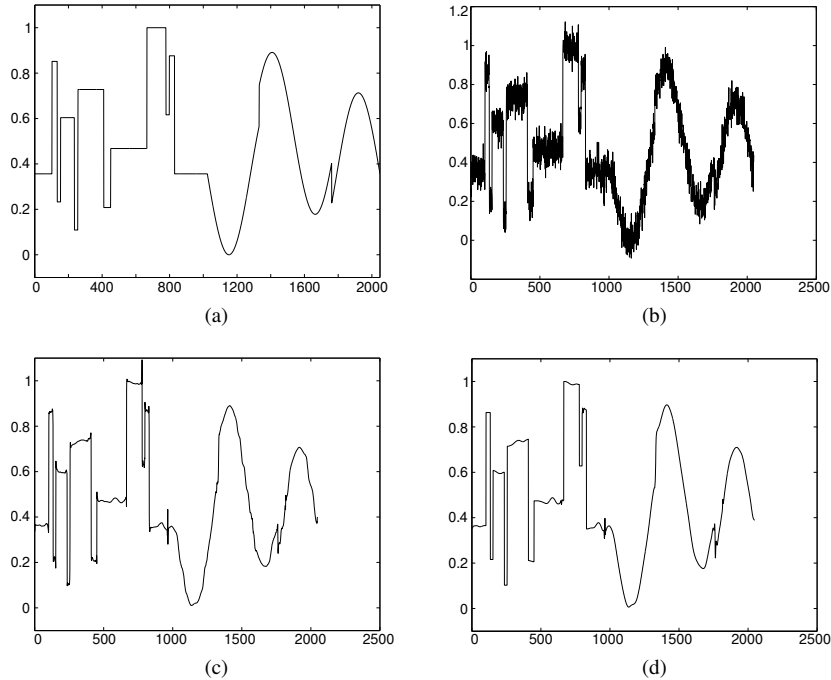


Fig. 13. *Blocks-HeaviSine* signal denoising, $\sigma_\epsilon = 5\%$ of total signal magnitude: (a) original signal, (b) noisy signal, (c) signal denoised using optimal wavelet transform (*Haar*) and (d) signal denoised using ICI-EPL.

be confirmed by visual inspection. The *Blocks* part of the denoised composite signal features almost perfect edge reconstruction, while the *HeaviSine* part of the signal is much smoother and more accurately reconstructed than in the *Haar* wavelet case, which points to a success of the adaptive algorithm.

5.3. Conclusion

In this paper, a point-wise adaptive lifting scheme for signal denoising was proposed. The goal was to develop an edge preserving transform, where edge is not considered to be only a step edge, but also any other sudden change in local signal properties. The adaptive algorithm selects appropriate wavelet basis, on each decomposition scale and for each signal sample, independently. In the lifting scheme P step, we use the statistical method of intersection of confidence intervals (ICI) rule to determine the neighborhood of samples which share the same local properties as the sample whose value is being predicted. This allows us to choose lifting filters which do not span across the edge between two regions with different local properties. In effect, longer and smoother wavelets are used for lower frequency regions, while shorter wavelets are used near edges in a signal.

To evaluate the efficiency of the proposed method, we used different signal classes and several levels of additive noise. In terms of the objective quality (RMSE), we showed that in most cases, the proposed method easily outperforms the original ICI smoothing method

and all the considered, commonly used, conventional wavelet transforms. When this was not the case, obtained results were comparable to the transform yielding the best result for a given signal. In terms of the subjective quality (visual quality) the difference was even more noticeable. Edge preserving nature of the transform could be clearly seen, as denoised signals did not exhibit edge oversmoothing nor usual unwanted visual artifacts, such as ringing.

Simulation results clearly showed that the proposed method performs very well about edges in a signal, which is often the crucial requirement in signal processing. As such, we see a great potential for its usage not only in signal denoising, but also as a preprocessing step in signal analysis applications.

References

- Claypoole, R., Davis, G., Sweldens, W., Baraniuk, R. (2003). Nonlinear wavelet transforms for image coding via lifting. *IEEE Transactions on Image Processing*, 12(12), 1449–1459. doi:10.1109/TIP.2003.817237.
- Daubechies, I., Sweldens, W. (1998). Factoring wavelet transforms into lifting steps. *Journal of Fourier Analysis and Applications*, 4(3) 245–267.
- Foi, A., Katkovnik, V., Egiazarian, K. (2007). Pointwise shape-adaptive dct for high-quality denoising and deblocking of grayscale and color images. *IEEE Transactions on Image Processing*, 16(5), 1395–1411. doi:10.1109/TIP.2007.891788.
- Katkovnik, V. (1999). A new method for varying adaptive bandwidth selection. *IEEE Transactions on Signal Processing*, 47(9) 2567–2571. doi:10.1109/78.782208.
- Katkovnik, V., Egiazarian, K., Astola, J. (2003). Adaptive varying scale methods in image processing. *Tampere International Center for Signal Processing, TICSP Series*, Vol. 19, Tampere, TTY, Monistamo.
- Katkovnik, V., Egiazarian, K., Astola, J. (2002). Adaptive window size image de-noising based on intersection of confidence intervals (ICI) rule. *Journal of Mathematical Imaging and Vision*, 16(3) 223–235. doi:10.1023/A:1020329726980.
- Kovacevic, J., Sweldens, W. (2000). Wavelet families of increasing order in arbitrary dimensions. *IEEE Transactions on Image Processing*, 9(3), 480–496. doi:10.1109/83.826784.
- Lerga, J., Susic, V., Vrankić, M. (2012). Separable image denoising based on the relative intersection of confidence intervals rule. *Informatica*, 22(3), 383–394. doi:10.1109/83.826784.
- Piella, G., Heijmans, H. (2002). Adaptive lifting schemes with perfect reconstruction. *IEEE Transactions on Signal Processing*, 50(7), 1620–1630. doi:10.1109/TSP.2002.1011203.
- Piella, G., Pesquet-Popescu, B., Heijmans, H. (2002). Adaptive update lifting with a decision rule based on derivative filters. *IEEE Signal Processing Letters* 9(10), 329–332. doi:10.1109/LSP.2002.804563.
- Sersic, D. (2000). Wavelet filter banks with adaptive number of zero moments. In: *Proc. 5th International Conference on Signal Processing WCCC-ICSP 2000*, Vol. 1, pp. 325–328. doi:10.1109/ICOSP.2000.894501.
- Stankovic, L. (2004). Performance analysis of the adaptive algorithm for bias-to-variance tradeoff. *IEEE Transactions on Signal Processing*, 52(5) 1228–1234. doi:10.1109/TSP.2004.826179.
- Su, C.-L. (2011). Ring line mapping and rotation inertial for fingerprint identification. *Informatica*, 22(3), 435–445.
- Sweldens, W. (1996). The lifting scheme: A custom-design construction of biorthogonal wavelets. *Applied and Computational Harmonic Analysis* 3(2) 186–200.
- Tomic, M., Sersic, D. (2012). Adaptive edge-preserving denoising by point-wise wavelet basis selection. *Signal Processing IET*, 6(1), 1–7. doi:10.1049/iet-spr.2010.0240.
- Tomic, M., Sersic, D., Loncaric, S. (2012). Adaptive spatio-temporal denoising of fluoroscopic X-ray sequences. *Biomedical Signal Processing and Control*, 7(2), 173–179. doi:10.1016/j.bspc.2011.02.003.
- Wu, Y., Pan, Q., Zhang, H., Zhang, S. (2004). Adaptive denoising based on lifting scheme. In: *Proc. 7th International Conference on Signal Processing ICSP'04*, Vol. 1, pp. 352–355. doi:10.1109/ICOSP.2004.1452654.
- Zhou, T.C., Zhou, H.M. (1999). *Adaptive eno-wavelet transforms for discontinuous functions*. CAM Report, No. 99-21, Department of Mathematics, UCLA 99, pp. 93–100.

M. Tomic is an assistant professor at the University of Rijeka, Faculty of Engineering. His main research interest include usage of wavelet transforms and statistical methods in signal denoising applications.

D. Sersic is an associate professor at the University of Zagreb, Faculty of Electrical Engineering and Computing, Croatia. He is currently with Colorado State University, Electrical and Computer Engineering Department, USA. His research interests are in field of adaptive wavelets, statistical estimation and modeling, blind separation and deconvolution.

Taškinės adaptyvios vilnelių transformacijos taikymas triukšmams iš signalų šalinti

Mladen TOMIC, Damir SERSIC

Nesugebėjimas pašalinti triukšmų iš signalo aukštų dažnių srities yra dažna triukšmų pašalinimo metodų problema. Nors vilnelių transformacijos yra šiek tiek efektyvesnės šiuo požiūriu nei kitos erdvinės ar dažninės transformacijos, tačiau ir čia susiduriama su sunkumais. Straipsnio autoriai triukšmams šalinti siūlo naudoti taškinę adaptyviąją vilnelių transformaciją. Trumpalaikėms signalo savybėms įvertinti panaudojama pasikliautinųjų intervalų sankirtos taisyklė. Todėl siūlomas metodas pasižymi efektyviu triukšmų pašalinimu iš aukštų dažnių srities, tuo pačiu neprarandant efektyvumo žemų dažnių srityje. Metodo efektyvumas buvo eksperimentiškai ištirtas su keletu signalų klasių. Gauti rezultatai patvirtino teorines prielaidas ir parodė siūlomojo metodo pranašumą prieš kitus triukšmų pašalinimo metodus.