

Robust L-Estimation Based Forms of Signal Transforms and Time-Frequency Representations

Igor Djurović, LJubiša Stanković, Johann F. Böhme

Abstract— The L -estimation based signal transforms and time-frequency (TF) representations are introduced by considering the corresponding minimization problems in the Huber estimation theory. The standard signal transforms follow as the maximum likelihood solutions for the Gaussian additive noise environment. For signals corrupted by an impulse noise the median based transforms produce robust estimates of the non-noisy signal transforms. When the input noise is a mixture of Gaussian and impulse noise, the L -estimation based signal transforms can outperform other estimates. In quadratic and higher order TF analysis the resulting noise is inherently a mixture of the Gaussian input noise and an impulse noise component. In this case, the L -estimation based signal representations can produce the best results. These transforms and TF representations give the standard and the median based forms as special cases. A procedure for parameter selection in the L -estimation is proposed. The theory is illustrated and checked numerically.

I. INTRODUCTION

Huber's estimation theory gives fundamental principles for solving a wide class of problems when the signal is influenced by impulse disturbances [1], [2]. Three groups of methods for signal parameters estimation are proposed. They are referred to as the maximum likelihood estimations (M -estimations), linear combination of order statistics estimations (L -estimations), and estimation derived from rank tests (R -estimations).

The standard Fourier transform can be obtained by solving a proper minimization problem with the squared absolute error as a loss function [3]-[6]. It is the maximum likelihood (ML) estimate of the nonnoisy signal transform for the Gaussian noise environment.

IEEE Transactions on Signal Processing, Vol.51, No.7, July 2003.

However, for an impulse kind of noise, the standard Fourier transform may produce poor results. Applying results from the Huber's estimation theory, Katkovnik has introduced the robust M -periodogram via the robust M -Fourier transform [3]. This method has been extended to the robust TF analysis of nonstationary signals embedded in an impulse noise [5], [7]. The robust M -periodogram definition is based on the absolute error as a loss function. Since this type of loss function does not produce a closed form solution, the iterative procedures are used in calculation of the robust M -periodogram [3]. Recently, it has been shown that a form of the robust Fourier transform can be obtained without iterative procedures, by using the median filter approach [8].

In this paper, we first have introduced the general form of discrete unitary transforms as a solution of an appropriate minimization problem. The standard form of these transforms follows as a solution of the minimization problem with the squared absolute error as a loss function. It is the ML estimate of the transform calculated for a signal in the Gaussian input noise. When the resulting signal is a sum of Gaussian and impulse noise (of a known type), we can calculate the ML estimate of a discrete unitary transform by solving an appropriate minimization problem with the corresponding loss function, defined by the resulting noise properties. However, this loss function is of the form which cannot be practically used. Huber's estimation theory provides solutions for this kind of problems [1]. They are based on the L and R -estimation approaches. These results are used here for a definition of the L -estimation forms of the discrete unitary transforms. These forms can

produce accurate results for a wide range of weights in the mixture of Gaussian and impulse noise. The well-known standard transforms and the recently introduced marginal median based form of these transforms easily follow as special cases from their L -estimation forms. A signal dependent adaptive procedure for determination of the parameter in the L -estimation based discrete unitary transforms is proposed. The presented theory is extended to the quadratic time-frequency (TF) representations as standard tools for analysis of non-stationary signals [9]-[12]. These representations can be interpreted as a Fourier transform of the signal local autocorrelation functions (LAF). As it will be shown in the paper, for quadratic TF distributions the resulting noise in the LAF can be treated as a mixture of the Gaussian and impulse noise, even for the pure Gaussian input noise. This suggests that the L -estimation forms can outperform the standard TF representation forms, even in some cases when the input noise is purely Gaussian. The presented L -estimation-based forms of signal transforms and representations can be used in a straightforward manner to define corresponding R -estimation-based forms. The considered signal transforms can be efficiently used in the robust filter design [24] and the parametric signal estimation.

The paper is organized as follows. Basic theory, along with definitions of the M - and L -estimates of the discrete unitary transforms, is given in Section II. The relationship between common estimation forms is derived. An extension of the L -estimation-based approach to the TF representations is given in Section III. Adaptive procedure for selection of the L -estimate parameter is presented in Section IV. In this section an analysis of statistical performance of the considered transforms is done, as well. Concluding remarks are given in Section V.

II. BASIC THEORY

Numerous discrete transforms have been defined for the analysis of discrete-time signals (DFT, Hadamard, discrete Walsh, Haar, discrete cosine, discrete sine, discrete Hartley transform, etc.). These transforms can be

written as:

$$S(k) = \sum_{n=0}^{N-1} s(n)\varphi_k(n), \quad (1)$$

where $s(n)$ is a signal of the length (or periodicity) N , and $\varphi_k(n)$, $k \in [0, N-1]$ are the basis functions. Without loss of generality we will consider a set of orthonormal basis functions

$$\sum_{k=0}^{N-1} \varphi_k(n)\varphi_k^*(n') = \delta(n-n'). \quad (2)$$

If (2) holds, $S(k)$ is a discrete unitary transform.

Consider the signal $s(n)$ embedded in a white noise $\nu(n)$, $x(n) = s(n) + \nu(n)$. Our goal is to estimate the signal transform coefficients $S(k)$ by using noisy samples $x(n)$. Estimated values will be denoted by $\hat{S}(k)$. One approach to solve this problem is based on the minimization of:

$$L_\varphi(m; k) = \sum_{n=0}^{N-1} F(Nx(n)\varphi_k(n) - m), \quad (3)$$

where $F(e)$ is a loss function, while

$$e(m; n, k) = Nx(n)\varphi_k(n) - m \quad (4)$$

is the error function. Relationship (3) can be understood as a redefinition of the discrete unitary transforms within the scale-location problem [2]. This problem also known as orthogonal robust regression has attracted significant attention in the statistical community. Details on this topic can be found in [1], [2], [13]-[17]. The minimum of (3) can be determined from

$$\frac{L_\varphi(m; k)}{\partial m^*} \Big|_{m=m_0} = 0, \text{ with } \hat{S}(k) = m_0. \quad (5)$$

Solution of (3) and (5) is called the M -estimate.

For the loss function $F(e) = |e|^2$ we easily get the standard discrete unitary transform:

$$\begin{aligned} X(k) &= \hat{S}(k) = \sum_{n=0}^{N-1} x(n)\varphi_k(n) = \\ &= \text{mean}\{Nx(n)\varphi_k(n) : n \in [0, N-1]\}. \end{aligned} \quad (6)$$

This solution has nice estimation properties when the noise is Gaussian. It can be shown that (6) is the ML estimate for this particular kind of noise. Namely, for a given probability density function (pdf) of noise, $p_\nu(e)$, the ML approach suggests the loss function $F(e) \sim -\log p_\nu(e)$ [3], [4]. For the Gaussian pdf the loss function assumes the form $F(e) = |e|^2$, producing (6) as the ML estimate. However, the ML estimates are quite sensitive to the variation of the noise pdf form. This means that the standard transforms, being the ML estimates for the Gaussian noise, may not produce satisfactory results for other kinds of noise.

This fact motivated introduction of the robust, instead of the ML estimates [1]. The robust estimate is introduced for a class of noises by taking the ML estimate of the worst noise from this class (noise with the longest tail) as the robust estimate for the whole class. The Laplacian noise with $F(e) = |e|$ is the worst case for numerous forms of impulse noises. Of course, this robust estimate based on $F(e) = |e|$ will produce worse results for the Gaussian noise than its ML estimate (6). However, the results for Gaussian noise will be only slightly worse, whereas the improvement of the estimation accuracy for impulse noise environment will be significant.

Solving (3) and (5) for the loss function $F(e) = |e|$ requires handling of nonlinear equations. We can distinguish two cases:

(a) The signal $x(n)$ and the basis functions $\varphi_k(n)$ are real-valued. Here the solution can be reduced to the median filter [8], [18] given as:

$$X_M(k) = \text{median}\{Nx(n)\varphi_k(n) : n \in [0, N-1]\}. \quad (7)$$

(b) The signal and/or the basis functions are complex-valued. Here, we will use the loss function $F(e) = |\text{Re}(e)| + |\text{Im}(e)|$ and marginal median approach [8], [18], that produces solution as:

$$\begin{aligned} X_M(k) = & \text{median}\{\text{Re}\{Nx(n)\varphi_k(n) : n \in [0, N-1]\}\} + \\ & j \text{median}\{\text{Im}\{Nx(n)\varphi_k(n) : n \in [0, N-1]\}\}. \end{aligned} \quad (8)$$

Note that the solution in this case can also be obtained by using the iterative procedure [3]-[6] or vector median approach; see Appendix A.

Signals are often corrupted not by a pure Gaussian or impulse noise but by their combination. In TF representations, due to their quadratic nature, resulting noise is a mixture of the Gaussian and impulse noise for the input noise being purely Gaussian. Theoretically, when we have a sum of Gaussian and impulse noise (of a known type) we can derive the pdf function, and its corresponding loss function for the ML estimation. However, this loss function will be of the form that is not practically applicable in the minimization. Huber's estimation theory provides solutions for this kind of problems [1]. They are based on the L - and R -estimation approaches. For real-valued signals and basis functions the L -estimation of transform coefficients is

$$X_L(k) = \sum_{i=0}^{N-1} a_i \mathbf{x}_{(i)}(k), \quad (9)$$

with $\sum_{i=0}^{N-1} a_i = 1$. The values $\mathbf{x}_{(i)}(k)$, $i = 0, \dots, N-1$, are elements from the set $\mathbf{E}_k = \{Nx(n)\varphi_k(n) : n \in [0, N-1]\}$, ordered into the nonincreasing sequence

$$\begin{aligned} \mathbf{x}_{(0)}(k) & \geq \dots \geq \mathbf{x}_{(i)}(k) \geq \\ \mathbf{x}_{(i+1)}(k) & \geq \dots \geq \mathbf{x}_{(N-1)}(k). \end{aligned} \quad (10)$$

Note that the standard discrete unitary transform, i.e., transforms based on the mean, and the transforms based on the median can be obtained as special cases of (9):

a) The standard discrete transforms (1) follows from (9) with $a_i = 1/N$, $i = 0, \dots, N-1$.

b) The robust median-based transforms result for

$$\begin{aligned} a_i & = \begin{cases} 1 & i = (N-1)/2 \\ 0 & i \neq (N-1)/2 \end{cases} \quad \text{for odd } N, \\ a_i & = \begin{cases} \frac{1}{2} & i \in [N/2-1, N/2] \\ 0 & \text{elsewhere} \end{cases} \quad \text{for even } N. \end{aligned} \quad (11)$$

Special attention will be paid to the form of coefficients in (9), which can be written in

analogy to the α -trimmed mean in the nonlinear digital filter theory [18]. The coefficients a_i , for even N , are given by

$$a_i = \frac{1}{N(1 - 2\alpha) + 4\alpha} \text{ for } i \in [(N - 2)\alpha, \alpha(2 - N) + N - 1] \quad (12)$$

and $a_i = 0$, elsewhere. From (12), the standard transforms follow for $\alpha = 0$, whereas $\alpha = 0.5$ produces the transforms based on the median filter. For $0 < \alpha < 0.5$, the transforms having performance between these two limit cases are obtained. For a mixture of Gaussian and impulse noise, the L -estimation with properly chosen value of α can produce more accurate results than either the standard or the median based transforms. This is especially important for quadratic and higher-order TF representations, where resulting noise inherently has a form of this mixture.

The L -estimation based filters (L -filters) and general Huber estimation theory have attracted significant attention in the signal and image filtering, including signal de-noising [18]-[23]. However, almost all of these filters produce low-pass characteristics. They are not suitable for signals with a high frequency content. The L -estimations of the discrete unitary transforms, proposed in this paper, are used for development of the robust filters in the frequency domain [24]. These filters can produce all filtering characteristics (lowpass, highpass, bandpass, and stopband). They are counterparts of the robust filters admitting negative weights recently proposed by Arce *et al.* [25]. An alternative approach for development of the robust filters of signals with high frequency content is presented in [21], where Schick and Krim created a wavelet-based de-noising method by applying data description length as a criterion for trade-off between ‘goodness-of-fit’ and model complexity. The probabilistic model used in [21] is inspired by the Huber’s work as well.

Now, we will illustrate these transforms on a real-valued transform example.

Example 1. Consider the Hadamard transform (HT) with the rectangular-shaped basis functions $\varphi_k(n)$ [26, pp.290-291], and $N = 128$. The signal $s(n)$ is formed as a sum of

two basis functions of the HT

$$s(n) = \varphi_{13}(n) + \varphi_{107}(n). \quad (13)$$

The standard HT of the nonnoisy signal (13) is:

$$S(k) = \sum_{n=0}^{N-1} s(n)\varphi_k(n) = \begin{cases} 1 & k = 13 \text{ or } k = 107 \\ 0 & \text{elsewhere.} \end{cases} \quad (14)$$

Signal is corrupted by a mixture of Gaussian and impulse noise:

$$x(n) = s(n) + \nu(n) = s(n) + \sigma_g \nu_1(n) + (a_h \nu_2(n))^3, \quad (15)$$

where $\nu_i(n)$, $i = 1, 2$, are mutually independent Gaussian white noises with unitary variances. A cube of the Gaussian noise is used hereafter as a model of impulse noise. Other models of impulse noise can be found in [27]-[29]. The HT estimations based on the mean, the α -trimmed mean with $\alpha = 3/8$, and the median are shown in Fig.1 for three different ratios of Gaussian and impulse noise. For a relatively small impulse component in the mixture (top row of Fig.1) the median-based form is worse than the mean, and the α -trimmed mean-based transforms. For $\sigma_g = 0.7$ and $a_h = 0.7$, all three transforms exhibit very similar performance (middle row of Fig.1). With an increase of the impulse noise component, $\sigma_g = 0.7$, $a_h = 1.0$, the standard HT becomes useless, whereas the median and the α -trimmed mean have very similar performance (bottom row of Fig.1). Note that typical realizations of the considered transforms are shown in Fig.1, whereas the statistical comparison is given in Section IV. From these three typical realizations we can conclude the following: a) In the Gaussian noise environment the α -trimmed-mean based HT performs as equally well as the standard mean-based transform; b) In the dominant impulse noise, the α -trimmed mean-based HT behaves as equally well as the median-based transform. Therefore, it can be a good choice for a whole variety of noise forms, from the pure Gaussian to the

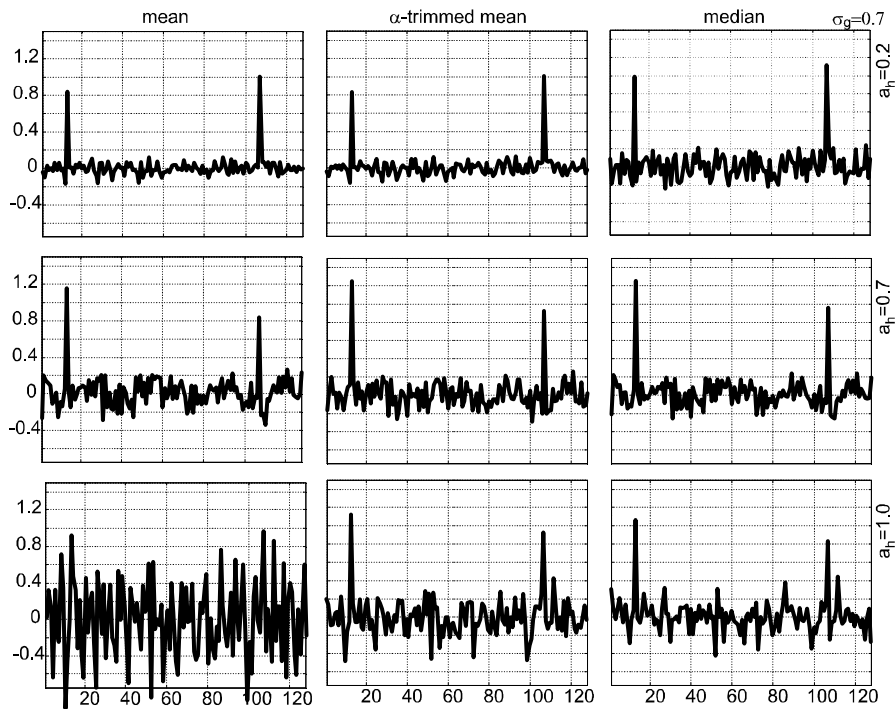


Fig. 1. HT of the signal corrupted by a mixture of the Gaussian and the impulse noise. Top row - ($\sigma_g = 0.7, a_h = 0.2$). Middle row - ($\sigma_g = 0.7, a_h = 0.7$). Bottom row - ($\sigma_g = 0.7, a_h = 1.0$).

pure impulse ones. A detailed statistical confirmation of this conclusion will be provided later.

Note: For the complex-valued signal and/or basis functions, the L -estimation based transform is given by

$$X_L(k) = \sum_{i=0}^{N-1} a_i \mathbf{r}_{(i)}(k) + j \sum_{i=0}^{N-1} a_i \mathbf{i}_{(i)}(k), \quad (16)$$

where $\mathbf{r}_{(i)}(k)$ and $\mathbf{i}_{(i)}(k)$ are elements belonging to the sets $\mathbf{R}_k = \{\text{Re}\{Nx(n)\varphi_k(n)\} : n \in [0, N-1]\}$, and $\mathbf{I}_k = \{\text{Im}\{Nx(n)\varphi_k(n)\} : n \in [0, N-1]\}$, respectively, sorted into the non-increasing sequences. Both the standard and the median-based transforms follow as special cases of (16), with a) $a_i = 1/N, i = 0, \dots, N-1$ and b) a_i given by (11), respectively. The L -estimation form in this case follows the marginal median definition, according to the facts presented in Appendix A.

III. TIME-FREQUENCY REPRESENTATIONS

TF representations are introduced for analysis of signals whose spectral content changes in time [9]-[12]. Influence of the Gaussian noise to the TF representations is a well studied topic [30]. In order to improve the performance of TF representations when the noise is not Gaussian, the L -estimation based definitions of these distributions will be introduced in this section. The L -estimation form of the short-time Fourier transform (STFT), as a time-varying form of the robust DFT, is defined first. Then, the L -estimation of the Wigner distribution (WD) is considered. Due to the nonlinearity of the WD, even for the signal with an additive Gaussian input noise, the resulting noise in this distribution is a combination of the Gaussian and impulse noise. The same holds for any other quadratic or higher-order TF representation.

A. *L-estimation of the DFT and STFT*

The DFT has complex-valued basis functions $\varphi_k(n) = \exp(-j2\pi nk/N)/\sqrt{N}$. The *L*-estimate of the DFT is obtained by using (16). The STFT is a time-varying form of the DFT. The *L*-estimate of the STFT can be defined by

$$STFT_L(n, k) = \text{Re}\{STFT_L(n, k)\} + j \text{Im}\{STFT_L(n, k)\} \\ = \sum_{i=0}^{N-1} a_i \mathbf{r}_{(i)}(n, k) + j \sum_{i=0}^{N-1} a_i \mathbf{i}_{(i)}(n, k), \quad (17)$$

where $\mathbf{r}_{(i)}(n, k)$ and $\mathbf{i}_{(i)}(n, k)$ are elements belonging to the sets $\mathbf{R}_{n,k} = \{\text{Re}\{\sqrt{N}x(n+m)\exp(-j2\pi km/N)\} : m \in [0, N-1]\}$, and $\mathbf{I}_{n,k} = \{\text{Im}\{\sqrt{N}x(n+m)\exp(-j2\pi km/N)\} : m \in [0, N-1]\}$, respectively, sorted into the nonincreasing sequences. For constant coefficients $a_i = 1/N, i = 0, 1, \dots, N-1$, in (17), the standard STFT follows. The median based STFT is obtained for a_i , given by (11). The spectrogram based on the *L*-estimation of the STFT is

$$SPEC_L(n, k) = |STFT_L(n, k)|^2 = |\text{Re}\{STFT_L(n, k)\}|^2 + |\text{Im}\{STFT_L(n, k)\}|^2. \quad (18)$$

B. *L-estimation of the Wigner distribution*

The WD is introduced in TF analysis in order to overcome low TF resolution of the STFT. It is defined as a DFT of the LAF $r_x(n, m) = x(n+m)x^*(n-m)$. From this fact the WD can be defined as a solution which minimizes the following functional:

$$L_{WD(n,k)} = \sum_{m=-N/2}^{N/2} F\left(\sqrt{N+1}x(n+m)x^*(n-m) \times \exp(-j4\pi km/(N+1)) - WD(n, k)\right). \quad (19)$$

The standard WD follows from (19) for the loss function $F(e) = |e|^2$. The median form of the robust WD is obtained with $F(e) = |e|$ [8], [18]. The *L*-estimate of the WD can be

introduced, according to the above analysis, as

$$WD_L(n, k) = \sum_{i=0}^N a_i \mathbf{x}_{(i)}(n, k), \quad (20)$$

where $\mathbf{x}_{(i)}(n, k)$ are the order statistics of

$$\mathbf{E}_{n,k} = \{\sqrt{N+1} \text{Re}\{x(n+m)x^*(n-m) \times \exp(-j4\pi km/(N+1))\} : m \in [-N/2, N/2]\}. \quad (21)$$

Special cases of the *L*-estimation based WD (20) are the standard and median WD. More details about the robust WD are given in [7], [8]; see Appendix B.

Next, we will show that the *L*-estimation based WD can outperform the standard WD, even in the case of the Gaussian input noise.

C. *ML estimation of the Wigner distribution*

Let the signal $s(n)$ be corrupted by an additive noise $\nu(n)$. The LAF is given by:

$$x(n+m)x^*(n-m) = s(n+m)s^*(n-m) + s(n+m)\nu^*(n-m) + \nu(n+m)s^*(n-m) + \nu(n+m)\nu^*(n-m) = r_s(n, m) + \Psi(n, m), \quad (22)$$

where $r_s(n, m) = s(n+m)s^*(n-m)$ is the signal component, while the noise influenced term is

$$\Psi(n, m) = s(n+m)\nu^*(n-m) + \nu(n+m)s^*(n-m) + \nu(n+m)\nu^*(n-m). \quad (23)$$

Assume that the input noise is of the form $\nu(n) = \nu_1(n) + j\nu_2(n)$, where $\nu_i(n), i = 1, 2$ are mutually independent white Gaussian noises $\mathcal{N}(0, \sigma^2)$. The component $s(n+m)\nu^*(n-m) + \nu(n+m)s^*(n-m)$ is a Gaussian white nonstationary noise with variance depending on the signal $s(n)$. The noise LAF $r_\nu(n, m) = \nu(n+m)\nu^*(n-m)$ can be written as

$$\nu(n+m)\nu^*(n-m) = \nu_1(n+m)\nu_1(n-m) + \nu_2(n+m)\nu_2(n-m) - j\nu_1(n+m)\nu_2(n-m) + j\nu_2(n+m)\nu_1(n-m). \quad (24)$$

For $m = 0$, it is equal to $r_\nu(n, 0) = \text{Re}\{\nu(n)\nu^*(n)\} = \nu_1^2(n) + \nu_2^2(n)$. The pdf of this noise is $p(\xi) = \exp(-\xi/2\sigma^2)/2\sigma^2$ for $\xi > 0$ and $p(\xi) = 0$ for $\xi < 0$ [31, eqs. (5.9) and (6.39)]. For $m \neq 0$ the real and imaginary part of noise $r_\nu(n, m) = r_1(n, m) + jr_2(n, m)$ in (24) can be written as $r_i(n, m) = \nu_a\nu_b + \nu_c\nu_d$, where ν_a, ν_b, ν_c , and ν_d are mutually independent Gaussian noises. Then, components $r_i(n, m) = \nu_a\nu_b + \nu_c\nu_d$ have the Laplacian pdf $p(\xi) = \exp(-|\xi|/\sigma^2)/2\sigma^2$, as it is shown in [32] and [33]. They are dominant with respect to $r_\nu(n, 0)$, since they exist for each $m \neq 0$.

Thus, we can conclude that for a Gaussian input noise the resulting noise in the WD $\Psi(n, m)$ is a mixture of the Gaussian and Laplacian impulse noise.

Example 2. We have considered the resulting WD noise $\Psi(n, m)$ for the signal $s(n) = \exp(j128\pi(n/(N+1))^2)$ corrupted by a Gaussian noise $\nu(n)$, with $N = 256$ samples. The standard WD of this signal is concentrated along the line $k = 128n/(N+1)$ in the TF plane. The cases with noise variances $\sigma^2 = (0.01)^2$ and $\sigma^2 = 0.64$ are considered. In both cases, the pdf of the resulting noise is obtained numerically. The value of $-\log(p(e))$ for resulting noise is shown in Fig.2. For small noise ($\sigma^2 = (0.01)^2$), we get that the ML estimation is very close to the standard WD, since the loss function is of the form $-\log(p(e)) = F(e) \simeq |e|^2$. In this case the Gaussian component $s(n+m)\nu^*(n-m) + \nu(n+m)s^*(n-m)$ dominates over $\nu(n+m)\nu^*(n-m)$ in the resulting WD noise $\Psi(n, m)$. In the second case ($\sigma^2 = 0.64$), the Laplacian component $r_\nu(n, m)$ is significant in $\Psi(n, m)$. Then, $F(e) \simeq |e|$ would produce estimation very close to the ML one. However, for small $|e|$, the ML loss function is different from $F(e) = |e|$. Thus, it can be concluded that the L -estimate of the WD will be an appropriate choice.

IV. PROCEDURE FOR SELECTION OF PARAMETER α

A procedure for automatic selection of the coefficient α in (12) will be discussed here. For simplicity, we will consider transform with a real-valued signal and basis functions. The α -trimmed mean based transform, with limit

cases (the standard and the median transform), can be written as:

$$X_\alpha(k) = S(k) + R_e(k) \quad (25)$$

where $S(k)$ is the standard transform of the signal without noise (1), whereas $R_e(k)$ includes the residual noise component and the component due to the nonlinearity of the α -trimmed mean. Assume that the signal transform can be represented with a very small number of non-zero values for $k \in \mathbf{K}$, where the set \mathbf{K} consists of all transform coefficients where $S(k) \neq 0$

$$X_\alpha(k) = \begin{cases} S(k) + R_e(k) & k \in \mathbf{K} \\ R_e(k) & k \notin \mathbf{K}. \end{cases} \quad (26)$$

This holds when the analyzed signal can be represented as a sum of the basis functions. For noncorrelated $S(k)$ and $R_e(k)$, when we can assume that $\sum_{k \in \mathbf{K}} S(k)R_e(k)/N \approx 0$, the mean squared value of $X_\alpha(k)$ is

$$\frac{1}{N} \sum_{k=0}^{N-1} |X_\alpha(k)|^2 \approx \frac{1}{N} \sum_{k \in \mathbf{K}} |S(k)|^2 + \frac{1}{N} \sum_{k=0}^{N-1} |R_e(k)|^2. \quad (27)$$

It can be concluded that the transform which produces minimal $\sum_{k=0}^{N-1} |X_\alpha(k)|^2/N$ will give the minimal mean squared error (MSE), i.e., the minimal component caused by the residual noise and nonlinearity $\sum_{k=0}^{N-1} |R_e(k)|^2/N$. Therefore, the α -trimmed mean transform from the considered set of transforms with different α that produces minimal $\sum_{k=0}^{N-1} |X_\alpha(k)|^2/N$ is close to those that produces minimal MSE. Our statistical analysis suggests that this criterion performs well for noisy cases for signal to noise ratios up to 20dB.

The presented simple procedure produces accurate results in a case of signals that can be represented by a relatively small number of non-zero transform coefficients, as it will be illustrated on examples. In general this procedure could not produce particularly accurate results. Then, more complicated algorithms should be employed, like, for example, those

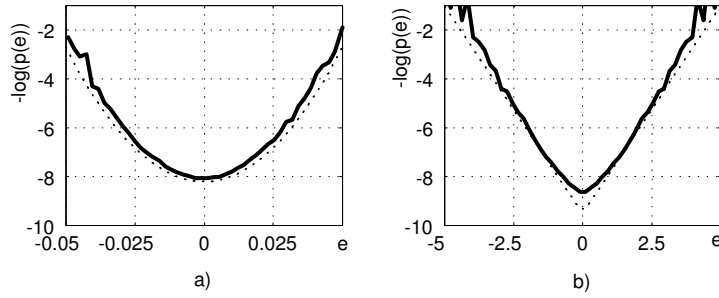


Fig. 2. Probability density function of the resulting noise $\Psi(n, m)$ for additive complex Gaussian noise. a) $\sigma = 0.01$. b) $\sigma = 0.8$.

presented in the case of the signal filtering in [34] and [35]. However, in opposite to the algorithms from [34] and [35], presented procedure is computationally very simple. It assumes calculation of several L -estimates that need only single sorting of the sequence for the considered coefficient. Remaining part of the procedure is evaluation of the very simple criterion.

Example 3. For statistical analysis we have considered the HT of signal (13) corrupted by a mixture of the Gaussian and impulse noise (15), where $\sigma_g \in [0, 1]$ and $a_h \in [0, 1.4]$. The MSE of the estimate is calculated as

$$MSE = \frac{1}{N} \sum_{k=0}^{N-1} |X_T(k) - S(k)|^2 \quad (28)$$

where $S(k)$ is given by (14), and $X_T(k)$ is the considered transform. The MSE for the standard HT, the α -trimmed mean for $\alpha = 3/8$ and $\alpha = 1/4$, and the median based HTs, for $\sigma_g = 0.7$ as a function of a_h is shown in Fig.3a. The MSE obtained with the proposed procedure is depicted with dotted line. The optimal L -estimate is chosen between the above mentioned transforms. From Fig.3a, it can be concluded that for small impulse noise (small a_h), the standard HT performs best, whereas with an increase of a_h , it becomes very poor (its MSE is a rapidly increasing function). The median based HT is relatively insensitive to the impulse noise (very slowly increasing function of a_h). The L -estimation based transforms follow the better of the previous two HT forms within the entire considered interval of the noise weights. In the interval $a_h \in [0.65, 1.1]$, they perform better than

either the mean or the median-based HT. Obviously, the adaptive transform performs statistically better than any transform with constant α .

Example 4. As a second example we have considered the DFT of signal:

$$s(n) = 16 \exp(-j2\pi n \cdot 13/256) + 16 \exp(-j2\pi n \cdot 107/256) \quad (29)$$

for $N = 256$, corrupted by a mixture of the Gaussian and impulse noise

$$\nu(n) = \sigma_g(\nu_1(n) + j\nu_2(n)) + (a_h\nu_3(n))^3 + j(a_h\nu_4(n))^3. \quad (30)$$

Note that the equivalent signal to noise ratio is the same as in the case of the HT since twice as many samples are taken. The MSE is shown in Fig.3b. The behavior of results, as well as the conclusions, are similar to those given for the HT.

Example 5. Finally, consider the WD of the signal $s(n) = 16 \exp(j128\pi(n/(N+1))^2)$, with $N = 256$ samples. The MSE as a function of a_h , for two fixed values $\sigma_g = 0.2$ and $\sigma_g = 0.5$, is shown in Fig.3c and d, respectively. In the first case, for a very small influence of the impulse noise, the standard WD performs best, but only slightly better than the α -trimmed mean for $\alpha = 3/8$. With an increase of the impulse noise, the standard WD becomes the worst. The L -estimations are better than the median based ones until the impulse noise influence becomes dominant $a_h > 0.9$. Even for $a_h > 0.9$, the L -estimation for $\alpha = 1/4$

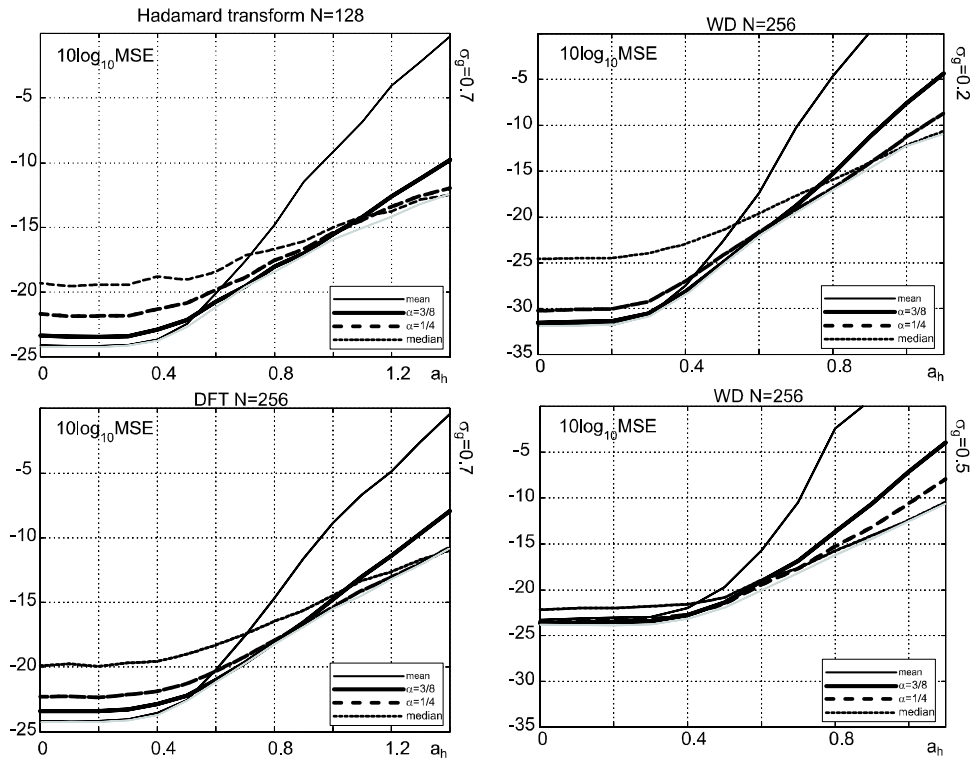


Fig. 3. MSE for transforms with constant and variable α . a) HT. b) DFT. c) WD for $\sigma_g = 0.2$. d) WD for $\sigma_g = 0.5$. Adaptive transforms are represented by the gray line.

is very close to the median. In the second case (Fig.3d), the standard WD is worse than the L -estimation based ones, even where there is no impulse component in input noise ($a_h = 0$). This is in accordance with the results derived in the previous section since the resulting noise in the WD is inherently a mixture of Gaussian and Laplacian noise. Other results are very similar to those in Fig.3c. In both cases the proposed procedure produces better accuracy than all WD forms with constant α .

Typical realizations of the standard WD, median-based WD, and the α -trimmed-based WD for $\alpha = 3/8$ are depicted with contour plots with five isolines in Fig.4 for $\sigma_g = 0.2$, $a_h = 0.6$. As expected from the statistical analysis (Fig.3c at $a_h = 0.6$), the α -trimmed based WD produces the best result with respect to the auto-term to noise ratio.

V. CONCLUSION

The L -estimation based discrete unitary transforms and TF representations are proposed as alternatives to their standard forms. They outperform the standard and median-based transforms and representations in the case of a mixture of Gaussian and impulse noise that inherently appears in quadratic and higher order TF representations. As a special case of the L -estimations, we have considered the α -trimmed mean transform. A simple procedure for adaptive selection of the parameter α in the α -trimmed mean transform is proposed. The adaptive transform can outperform all constant parameter signal transforms in the case of noisy signals. The L -estimations of the discrete unitary transforms proposed in this paper are used for development of the robust filters in the frequency domain [24]. These filters can produce all filtering characteristics (lowpass, highpass, band-

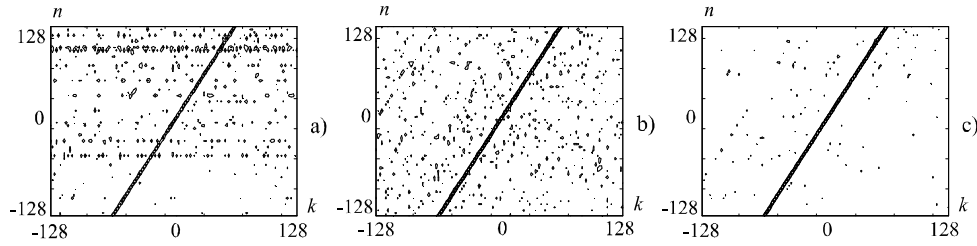


Fig. 4. WD of noisy signal. a) Standard WD. b) Median WD. c) α -trimmed mean WD. Noise is a mixture of Gaussian and impulse noise with $\sigma_g = 0.2$ and $a_h = 0.6$.

pass, and stopband). Another promising application field of the transforms proposed in this paper could be in the parametric estimation of the FM signals corrupted by the impulse noise.

Appendix A

When complex-valued functions are considered, the standard median cannot be used for definition of robust transforms. Instead of the standard median, iterative procedures or median forms (vector median and marginal median) derived for vector-valued functions must be used.

Iterative procedure: When the loss function $F(e) = |e|$ is used in the case of complex-valued signals, there is a problem to estimate the real and imaginary part of a transform. For the robust M -periodogram (squared absolute value of the robust Fourier transform) calculation, Katkovnik has proposed an iterative procedure [3], [4]. From the minimization problem (3) and (5), for the loss function $F(e) = |e|$, the nonlinear equation

$$X_I(k) = \gamma(k) \sum_{n=0}^{N-1} \frac{Nx(n)\varphi_k(n)}{|e(n, k)|} \quad (31)$$

follows, where $X_I(k)$ is the robust transform, and $e(n, k) = Nx(n)\varphi_k(n) - X(k)$, $\gamma(k) = 1/\sum_{n=0}^{N-1} 1/|e(n, k)|$. Since the unknown robust transform $X_I(k)$ is on both sides of (31), an appropriate iterative procedure should be used. Note that the nonlinear relation (31) is of the form $x = f(x)$, which can be solved by using the iterative method $x_{n+1} = f(x_n)$ with

a suitable initial guess x_0 . The standard transform (6) is used as the initial guess $X^{(0)}(k)$. Then the iterative procedure is applied as

$$\mathbf{X}^{(l)} = \mathbf{A}^{(l-1)}\mathbf{x} \quad (32)$$

where

$$\mathbf{X}^{(l)} = [X_I^{(l)}(0) \dots X_I^{(l)}(N-1)]^T \quad (33)$$

$\mathbf{A}^{(l-1)}$ is the matrix whose elements are

$$a(n, k) = N\gamma^{(l-1)}(k)\varphi_k(n)/|e^{(l-1)}(n, k)|, \quad (34)$$

with

$$e^{(l-1)}(n, k) = Nx(n)\varphi_k(n) - X_I^{(l-1)}(k) \quad (35)$$

$$\gamma^{(l-1)}(k) = 1/\sum_{n=0}^{N-1} 1/|e^{(l-1)}(n, k)| \quad (36)$$

and $\mathbf{x} = [x(0) \ x(1) \ \dots \ x(N-1)]^T$. The iterative procedure is stopped when

$$\max \left\{ |\mathbf{X}^{(l)} - \mathbf{X}^{(l-1)}|/|\mathbf{X}^{(l-1)}| \right\} \leq \varepsilon \quad (37)$$

where ε is a given precision. Realization of the similar iterative procedures, along with proof of their convergence, is discussed in [1, Sec.7.8], [3] and [36].

Note: Similar procedures can be used not only for the loss function $F(e) = |e|$ but for any other loss functions, like, for example, for the loss function that produces myriad filter [36], [37], whose form is $F(e) = \ln(|e|^2 + K^2)$. The basic difference is only in the matrix $\mathbf{A}^{(l)}$ coefficients.

Vector median: The minimization problem (3) and (5) can be solved for the complex-valued functions by using the vector median

approach [8], [38], [39]. If we restrict our solution to the set $\mathbf{E}_k = \{Nx(n)\varphi_k(n) : n \in [0, N-1]\}$, then the solution of the minimization problem with the loss function $F(e) = |e|$ is $X_V(k) = m(k) \in \mathbf{E}_k$ such that

$$\sum_{n=0}^{N-1} |m(k) - Nx(n)\varphi_k(n)| \leq \sum_{n=0}^{N-1} |Nx(n_1)\varphi_k(n_1) - Nx(n)\varphi_k(n)| \quad (38)$$

where $n_1 \in [0, N-1]$. This is a definition of the vector median. It interprets $|Nx(n_1)\varphi_k(n_1) - Nx(n)\varphi_k(n)|$ as a distance between two points $(Nx(n_1)\varphi_k(n_1))$ and $(Nx(n)\varphi_k(n))$ in the complex plane. The vector median is then the point $m(k) \in \mathbf{E}_k$ such that its sum of distances to all other points from the set \mathbf{E}_k is minimal. Therefore, the robust transform for complex-valued signals and/or basis functions can be defined as

$$X_V(k) =$$

$$\text{vector_median}\{Nx(n)\varphi_k(n) : n \in [0, N-1]\}. \quad (39)$$

Marginal median: Significant computational simplification of the previous iterative and vector median procedures can be achieved by using a separate minimization of the error function real and imaginary parts. If we choose the loss function as $F(e) = |\text{Re}(e)| + |\text{Im}(e)|$, with the assumption that real and imaginary part of the error function are statistically independent, the minimization problem is reduced to the real-valued case. Then, the robust transform is of the form [7]

$$X_M(k) =$$

$$\text{median}\{\text{Re}\{Nx(n)\varphi_k(n) : n \in [0, N-1]\}\} + j\text{median}\{\text{Im}\{Nx(n)\varphi_k(n) : n \in [0, N-1]\}\}. \quad (40)$$

This solution corresponds to the marginal median in the theory of filters [39]. From our statistical experiments, we concluded that the results obtained by using the previous two median forms and iterative procedure are very close [8]. In the mathematical literature, the

median of a complex-valued sequence [40] is sometimes defined as the marginal median.

Appendix B

The robust WD can be derived from the minimization problem (19), with the loss function $F(e) = |e|$. It assumes the form [7]

$$WD_I(n, k) = \gamma(n, k) \sum_{m=-N/2}^{N/2} \frac{\sqrt{N+1}}{|e(n, m, k)|} \times$$

$$x(n+m)x^*(n-m)e^{-j4\pi km/(N+1)} \quad (41)$$

with

$$e(n, m, k) = \sqrt{N+1}x(n+m)x^*(n-m) \times$$

$$\exp(-j4\pi km/(N+1)) - WD_I(n, k)$$

and $\gamma(n, k) = 1/\sum_{m=-N/2}^{N/2} 1/|e(n, m, k)|$. Equation (41) is an implicit definition of $WD_I(n, k)$. Thus, it is necessary to use the iterative procedure for the robust WD calculation [7]. In [7], it has been shown that the robust WD is real-valued. The property of realness holds for each iteration in the procedure. The marginal median-based WD then reduces to [7]

$$WD_M(n, k) =$$

$$\text{median}\{\sqrt{N+1} \text{Re}\{x(n+m)x^*(n-m) \times \exp(-j4\pi km/(N+1))\} : m \in [-N/2, N/2]\}. \quad (42)$$

VI. ACKNOWLEDGMENT

The authors are thankful to the Associate Editor and the reviewers for very valuable comments and references that helped to improve the paper. This work is supported by the Volkswagen Stiftung, Federal Republic of Germany. A part of the Igor Djurović's work is supported by the Postdoctoral Fellowship for Foreign Researchers of Japan Society for the Promotion of Science.

REFERENCES

- [1] P.J. Huber: *Robust statistics*, John Wiley&Sons Inc., 1981.
- [2] P. J. Huber: "Robust regression: asymptotics, conjecture and Monte Carlo," *Ann. Math. Stat.*, Vol. 1, No. 5, 1973, pp. 799-821.
- [3] V. Katkovnik: "Robust M-periodogram," *IEEE Trans. Sig. Proc.*, Vol. 46, No. 11, Nov. 1998, pp. 3104-3109.
- [4] V. Katkovnik: "Robust M-estimates of the frequency and amplitude of a complex-valued harmonic," *Sig. Proc.*, Vol. 77, No. 1, Aug. 1999, pp. 71-84.
- [5] L.J. Stanković, I.Djurović, S. Stanković: "The robust Wigner distribution," in *Proc. of IEEE ICASSP'2000*, Istanbul, Vol. I, pp. 77-80.
- [6] V. Katkovnik, I. Djurović, L.J. Stanković: "Instantaneous frequency estimation using robust spectrogram with varying window length," *AEU - Jour. Elec. Com.*, Vol. 54, No. 4, 2000, pp. 193-202.
- [7] I. Djurović, L.J. Stanković: "Robust Wigner distribution with application to the instantaneous frequency estimation," *IEEE Trans. Sig. Proc.*, Vol. 49, No. 12, Dec. 2001, pp. 2985-2993.
- [8] I. Djurović, V. Katkovnik, L.J. Stanković: "Median filter based realizations of the robust time-frequency distributions," *Sig. Proc.*, Vol. 81, No. 7, July 2001, pp. 1771-1776.
- [9] L. Cohen: *Time-frequency analysis*, Prentice-Hall, 1995.
- [10] F. Hlawatsch, G.F. Boudreaux-Bartels: "Linear and quadratic time-frequency signal representations," *IEEE Sig. Proc. Mag.*, Apr. 1992, pp. 21-67.
- [11] F. Auger, P. Flandrin: "Improving the readability of time-frequency and time-scale representations by reassignment method," *IEEE Trans. Sig. Proc.*, Vol. 43, May 1995, pp. 1068-1089.
- [12] H.I. Choi, W.J. Williams: "Improved time-frequency representation of multicomponent signals using exponential kernels," *IEEE Trans. ASSP*, Vol. 73, No. 6, June 1989, pp. 862-871.
- [13] D.F. Andrews, P.J. Bickel, F.R. Hampel, P.J. Huber, W.H. Rogers, J.W. Tukey: *Robust Estimates of Location: Survey and Advances*, Princeton University Press, Princeton, N.J, 1972.
- [14] P.J. Huber: "Robust estimation of a location parameter," *Ann. Math. Statist.*, Vol. 35, 1964, pp. 73-101.
- [15] R.D. Martin, R.H. Zamar: "Efficiency-constrained bias-robust estimation of location," *Ann. Statist.*, Vol. 21, 1993, pp. 338-354.
- [16] D.M. Rocke, G.W. Downs: "Estimating the variances of robust estimators of location: influence curve, jackknife and bootstrap," *Comm. Stat. B.*, Vol. 10, 1981, pp. 221-248.
- [17] D.P. Wiens: "Minimax-variance L- and R-estimators of location," *Can. J. Statist.*, Vol. 18, No. 1, 1990, pp. 47-57.
- [18] I. Pitas, A.N. Venetsanopoulos: *Nonlinear digital filters: Principles and applications*, Kluwer, 1990.
- [19] M. Pappas, I. Pitas: "Multichannel distance filter," *IEEE Trans. Sig. Proc.*, Vol. 47, No. 12, Dec. 1999, pp. 3412-3416.
- [20] A. Ben Hamza, H. Krim: "Image denoising: A nonlinear robust statistical approach," *IEEE Trans. Sig. Proc.*, Vol. 49, No. 12, Dec. 2001, pp. 3045-3054.
- [21] H. Krim, I.C. Schick: "Minimax description length for signal denoising and optimized representation," *IEEE Trans. Inf. Th.*, Vol. 45, No. 3, Apr. 1999, pp. 898-908.
- [22] S. Burley, M. Darnell: "Robust impulse noise suppression using adaptive wavelet de-noising," in *Proc. of IEEE ICASSP'97*, Vol. 5, pp. 3417-3420.
- [23] P.L. Ainsleigh, C.K. Chui: "A B-wavelet based noise reduction algorithm," *IEEE Trans. Sig. Proc.*, Vol.44, No.5, May 1996, pp. 1279-1284.
- [24] I. Djurović, L.J. Stanković: "Realization of robust filters in the frequency domain," *IEEE Sig. Proc. Let.*, Vol. 9, No.10, Oct. 2002, pp. 333-335.
- [25] S. Kalluri and G.R. Arce: "Robust frequency-selective filtering using weighted myriad filters admitting real-valued weights," *IEEE Trans. Sig. Proc.*, Vol. 49, No. 11, Nov. 2001, pp. 2721-2733.
- [26] K.R. Castleman: *Digital image processing*, Prentice Hall 1996.
- [27] S.A. Kassam: *Signal detection in Non-Gaussian noise*, Springer-Verlang, 1988.
- [28] D. Middleton: "Non-Gaussian noise models in signal processing for telecommunications: New methods and results for class A and class B noise models," *IEEE Trans. Inf. Th.*, Vol. 45, No. 4, May 1999, pp. 1129-1150.
- [29] C.L. Nikias, M. Shao: *Signal processing with alpha stable distributions and applications*, New York: Wiley 1995.
- [30] M.G. Amin: "Minimum variance time-frequency distribution kernels for signals in additive noise," *IEEE Trans. Sig. Proc.*, Vol. 44, No. 9, Sept. 1996, pp. 2352-2356.
- [31] A. Papoulis: *Probability, random variables and stochastic processes*, McGraw Hill, 1984.
- [32] W.J. Szajnowski: "Generation of a discrete-time correlated Laplacian process," *IEEE Sig. Proc. Let.*, Vol. 7, No. 3, Mar. 2000, pp. 69-70.
- [33] I. Djurović, L.J. Stanković, J. F. Böhme: "Estimates of the Wigner distribution in Gaussian noise environment," *AEU Int. Jour. El. Com.*, Vol. 56, No. 5, 2002, pp. 337-340.
- [34] C. Kotropoulos, I. Pitas, "Constrained adaptive LMS L-filters," *Sig. Proc.*, Vol. 26, 1992, pp. 335-358.
- [35] T. Chen, H. R. Wu, "Recursive LMS L-filters for noise removal in images," *IEEE Sig. Proc. Let.*, Vol. 8, No. 2, Feb. 2001, pp. 36-38.
- [36] S. Kalluri, G.R. Arce: "Fast algorithms for weighted myriad computation by fixed-point search," *IEEE Trans. Sig. Proc.*, Vol. 48, No. 1, Jan. 2000, pp. 159-171.
- [37] I. Djurović, L.J. Stanković, J. F. Böhme: "Myriad filter based form of the DFT," in *Proc. of EUSIPCO 2002*, France, Toulouse, Sep. 2002, Vol.III, pp. 433-436.
- [38] J. Astola, P. Haavisto, Y. Neuvo: "Vector median filters," *Proc. IEEE*, Vol. 78, No. 4, Apr. 1990, pp. 678-689.
- [39] I. Pitas, P. Tsakalides: "Multivariate ordering in color image filtering," *IEEE Trans. Circ. Syst. for Video Tech.*, Vol. 1, No. 3, Sep. 1991, pp. 247-259.
- [40] A. Napolitano, C.M. Spooner: "Median-based cyclic polyspectra estimation," *IEEE Trans. Sig. Proc.*, Vol. 48, No. 5, May 2000, pp. 1462-1466.