

A Spatially Adaptive Nonparametric Regression Image Deblurring

Vladimir Katkovnik, Karen Egiazarian, *Senior Member, IEEE*, and Jaakko Astola, *Fellow, IEEE*

Abstract—We propose a novel nonparametric regression method for deblurring noisy images. The method is based on the local polynomial approximation (LPA) of the image and the paradigm of intersecting confidence intervals (ICI) that is applied to define the adaptive varying scales (window sizes) of the LPA estimators. The LPA-ICI algorithm is nonlinear and spatially adaptive with respect to smoothness and irregularities of the image corrupted by additive noise. Multiresolution wavelet algorithms produce estimates which are combined from different scale projections. In contrast to them, the proposed ICI algorithm gives a varying scale adaptive estimate defining a single best scale for each pixel. In the new algorithm, the actual filtering is performed in signal domain while frequency domain Fourier transform operations are applied only for calculation of convolutions. The regularized inverse and Wiener inverse filters serve as deblurring operators used jointly with the LPA-design directional kernel filters. Experiments demonstrate the state-of-art performance of the new estimators which visually and quantitatively outperform some of the best existing methods.

Index Terms—Adaptive scale, adaptive window size, deblurring, directional local polynomial approximation (LPA), nonparametric regression.

I. INTRODUCTION

SUPPOSE that we wish to reconstruct a two-dimensional (2-D) image intensity function y , but we are able to observe only $(y \otimes v)(x)$, where v is the point spread function (PSF). The blurring phenomenon, modeled by the kernel v (continuous or discrete) is very evident in many image processing applications. Such linear inverse problems arise in optical systems, satellite imaging, radiometry, ultrasonic and magnetic resonance imaging, etc.

We assume that the convolution is discrete and noisy so that we observe z given by

$$z(x) = (y \otimes v)(x) + \eta(x) \quad (1)$$

where x is defined on the regular $n_1 \times n_2$ integer lattice, $x \in X = \{(k_1, k_2), k_1 = 1, 2, \dots, n_1, k_2 = 1, 2, \dots, n_2\}$, and η is zero mean white Gaussian noise with the variance σ^2 . In the 2-D discrete Fourier transform (DFT) domain, we have for (1)

$$Z(f) = Y(f)V(f) + \eta(f) \quad (2)$$

with $Z(f)$, $Y(f)$, $V(f)$, and $\eta(f)$ of the respective size $n_1 \times n_2$ being the DFTs of the corresponding signals and $f \in F$, $F =$

$\{(f_1, f_2), f_i = 2\pi k_i/n_i, k_i = 0, 1, \dots, n_i - 1, i = 1, 2\}$ is the 2-D normalized frequency. Equation (2) means that the convolution (1) is circular.

It is assumed that v is known. In principle, an unbiased estimate of $Y(f)$ can be obtained as a straightforward pure (“naive”) inverse solution of the (2), i.e., $\hat{Y}(f) = Z(f)/V(f)$. However, in the cases of interest, v is not invertible, in the sense that the inverse transform does not exist as a bounded linear operator; such inverse problems are called ill posed or ill conditioned [23]. In the frequency domain, being ill posed means that $V(f)$ takes zero or close to zero values. If the system is ill conditioned, then the random component is enormously amplified in the naive inversion to yield an extremely noisy and useless estimate. It is now a standard to approach the inverse problem by the method of regularization, in which one applies, rather than the inversion, a regularized inverse operator [23]. It is defined by $\hat{Y}(f) = Z(f)V(-f)/(|V(f)|^2 + \varepsilon^2)$, where $\varepsilon > 0$ is a regularization parameter.

One of the typical approaches to the image reconstruction is based on projection methods assuming that the observed image can be decomposed as a sum of basis functions. A natural way of projection for ill-posed problems is associated with the singular value decomposition (SVD). Denote v^* the adjoint of the PSF v and assume that v^*v is an operator with positive eigenvalues and some eigenfunctions. These eigenfunctions are used as a basis in SVD methods (e.g., [2], [3], and references herein). Despite the great popularity of schemes based explicitly or implicitly on SVD, the method suffers from natural performance limitations. These are rooted in the fact that the basis eigenfunctions are derived from the convolution operator v and not from the object y . Thus, efficient representation of the object y by the eigenfunctions of v^*v cannot be guaranteed for all applications [2]. There are important alternative strategies based on developing basis functions for the object y rather than for the convolution operator v .

Wavelets provide effective solutions in terms of this approach. Many real-world images and signals have economical representations in which few large wavelet coefficients capture most of the signal energy. Signal and image reconstruction using wavelet shrinkage is a spatially adaptive process quite suited to signals with edges and other singularities [1], [5], [12], [20].

The wavelet-vaguelette decomposition has been proposed for the deconvolution problem in [4]. In this approach, the wavelet expansion is applied to the signal obtained as an output of the pure deconvolution. A scale-dependent shrinkage is employed to estimate the signal wavelet coefficients. Further development of the wavelet approach is proposed in [2] with a new

Manuscript received January 1, 2003; revised July 19, 2004. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Robert D. (G. E.) Nowak.

The authors are with the Signal Processing Laboratory, Tampere University of Technology, FIN-33101 Tampere, Finland (e-mail: katkov@cs.tut.fi; karen@cs.tut.fi; jta@cs.tut.fi).

Digital Object Identifier 10.1109/TIP.2005.851705

decomposition which is much better adapted to the type of edge-dominated objects while providing an almost diagonal representation of the Radon convolution operator. It is done by introducing new basis functions, so-called curvelets: smooth, anisotropic directional elements well suited for detecting and synthesizing curved edges. A new efficient hybrid wavelet based deconvolution algorithm was developed in [21]. This algorithm comprises Fourier-domain regularized inversion followed by wavelet-domain noise suppression. It is shown that the developed algorithm outperforms the conventional Wiener filters as well as some other wavelet-based image restoration algorithms. The expectation-maximization (EM) algorithm for image deblurring, based on penalized maximum likelihood setting, is proposed in [7]. The algorithm uses the EM technique in the wavelet domain. It is shown that the algorithm is very efficient and competitive.

A common point to all of the above methods that start from the frequency domain (2) is that the basis functions are applied for parametric approximations of the object function y in the form of a series with coefficients (parameters) defined from the observations. These functions may be Fourier harmonics in DFT, eigenfunctions of the convolution operator in SVD methods, or wavelets in wavelet based decompositions. There exist many deconvolution techniques based on this type of approaches.

Fundamentally different ideas and methods arise from a nonparametric regression approach [6], [13], [19]. The main goal of the nonparametric estimation is to build a pointwise approximation in the spatial domain using observations from some neighborhood. The estimation is produced in a sliding window what means that the every point of the object is treated in a special way in order to achieve the best possible performance. There is a number of proposals for nonparametric estimation of non-blurred noisy images which allow to preserve the sharp edge structure as well as the edge detection and reconstruction (e.g., [8], [15], [16], and [22]).

The adaptation used in this paper is based on pointwise nonparametric estimation of the image intensity function. The estimator is derived from the local polynomial approximation (LPA) of the image in a sliding window. The nonlinearity of the method is incorporated in an adaptive choice of the estimation window size [intersecting confidence intervals (ICI) rule], allowing a high-quality image and edge recovery [10], [14]–[16], [22]. The basic idea of this adaptation rule is quite simple. The algorithm searches for a largest local vicinity of the point of estimation where the LPA assumption fits well to the data. The estimates of the image are calculated for a grid of scales (window sizes) and compared. The adaptive window size is defined as the largest of those in the grid where the estimate does not differ significantly from the estimators corresponding to smaller window sizes [10], [18]. It is proved (see cited papers) that this adaptation results in estimates that are optimal in minimax sense for different classes of functions.

Overall, in terms of optimal asymptotic properties, these nonparametric regression estimates and the wavelet techniques are mainly equivalent. However, in practical aspects, the wavelets and the nonparametric regression methods are quite different. In particular, the standard application of the multiresolution

wavelet denoising includes the following three steps: wavelet analysis, thresholding (filtering) in transform domain, and signal synthesis (inversion of the wavelet transform). As a result, the estimate for each pixel is composed from the estimates of different scales. In the considered nonparametric regression approach, the best scale is selected for each pixel and this single scale estimate is used in the estimation. Thus, we arrive to the varying adaptive scale estimation with the single scale for each pixel.

The first application of this kind of adaptive nonparametric regression estimators for blurred one-dimensional (1-D) observations has been reported in [9], where the window size adaptive kernel estimate with the ICI rule is applied to a signal obtained by the pure deconvolution. The continuous time model is used for the observation model, the algorithm presentation as well as for the analysis. The accuracy analysis shows that the estimator possesses the best possible ability for the pointwise adaptive estimation.

In this paper, the approach proposed in [9] is extended to 2-D discrete data imaging, as well as to the biased regularized and Wiener inverse estimators. The remainder of this paper is organized as follows. In Section II, the proposed adaptive algorithm is presented. Some analytical results concerning the asymptotic analysis are shown in Section III. Experiments with test images and comparison with competitive algorithms are discussed in Section IV.

II. ADAPTIVE DEBLURRING

A. LPA Kernel Estimates

Let us consider a linear discrete kernel estimator defined on the integer lattice X and given by the kernel $g_h(x)$, $x \in X$, with the scale (window size) parameter $h > 0$

$$y_h(x) = (g_h \otimes y)(x). \quad (3)$$

Thus, y_h is an output of the filter defined by g_h and \otimes stays for the 2-D discrete convolution.

The following is assumed for g_h .

- 1) The m vanishing moments: $(g_h \otimes x^k)(0) = \delta_{|k|,0}$, $|k| \leq m$, where $k = (k_1, k_2)$ is a multi-index, k_1, k_2 are non-negative integers, and $|k| = k_1 + k_2$, $x^k = x_1^{k_1} x_2^{k_2}$, $\delta_{|k|,0}$ is equal to one for $|k| = 0$ and to zero otherwise.
- 2) $\|g_h\|^2 = \sum_x |g_h(x)|^2 \leq B h^{-b}$, $B, b > 0$.

We say that the g_h is a *smoothing kernel estimator* (low-pass filter) of the *order* m . The order means that the operator is reproducing with respect to 2-D polynomials of the degree less than or equal to m . If $y(x)$ is a such polynomial then $(g_h \otimes y)(x) = y(x)$. The condition 2) means that the g_h is a low-pass filter. Its passband becomes narrower when the scale h increases.

If h is small, the passband of the estimator is large and the signal $y(x)$ is reproduced without distortions. If h is large, then the passband of the estimate is narrow and only low frequency components of y can be observed at the output of the estimator. Thus, for large h , high-frequency features of y are smoothed (suppressed) and the output y_h is smoother than y . For small h , all the high-frequency features and details of y are preserved in y_h . If the input signal y is given with an additive random

noise, then there exists an optimal value of h corresponding to the usual tradeoff between random and systematic errors of the signal reconstruction. The index h in notation y_h emphasizes that we deal with the smoothed version of the signal y and the involved level of smoothing is defined by h .

The design of g_h with the properties 1) and 2) can be produced by the LPA method that can be found in [6], [13], [15], [16], and [19]. The concept of the LPA is quite natural. It is assumed that the object function y is well approximated by a polynomial in a neighborhood of the point of interest x . We find the coefficients of the polynomial fit by the weighted least-square method and use this approximation in order to calculate the estimate for the point of interest x , called also a “center” of the LPA. In fact, the local expansion is applied in order to calculate the estimate for this point of interest only. For another point, the calculations are repeated. This pointwise procedure determines a nonparametric character of the LPA estimation.

In this paper, we are mainly concerned with a special case of the LPA when the estimate can be presented as the convolution of 2-D image and 2-D invariant mask (kernel) defined on the regular integer lattice. Then the LPA estimate y_h takes a form (3) with the shift invariant kernel g_h [15], [16]

$$g_h(x) = w_h(x)\phi^T(0)\Phi^{-1}\phi(x) \\ \Phi = \sum_x w_h(x)\phi(x)\phi^T(x) \quad (4)$$

where $\phi(x)$ is a vector of linear independent 2-D polynomials of the degree from 0 up to m and the window function $w_h(x) = w(x/h)/h^2$ is used to formulate mathematically the fitting localized in a neighborhood of the center x . The scale parameter $h > 0$ determines the size of the neighborhood. With quite natural assumptions on the window function w , the kernel (4) gives a filter which satisfies both of the conditions 1) and 2).

We wish to note that actually the LPA has appeared in signal processing in a number of modifications and under different names: sliding (moving) least square, weighted least square, Savitzky–Golay filter, moment filters, etc. We prefer an established and explicitly clear term LPA with references to publications on nonparametric regression in statistics. Note also that the LPA is one of the most popular forms of modeling nonparametric regression phenomena.

B. Idea of Deblurring Algorithm

A basic idea of the proposed deblurring algorithm is to use the smoothed image intensity y_h instead of the original y as a solution of the inverse problem and to exploit the kernel estimator g_h equipped with the scale parameter h in order to suppress the noise as much as possible while preserving details of the object function y . Applying the kernel operator g_h to the both sides of (1) yields

$$z_h(x) = (g_h \otimes (y \otimes v))(x) + \eta_h(x) \\ = ((g_h \otimes y) \otimes v)(x) + \eta_h(x) = (y_h \otimes v)(x) + \eta_h(x). \quad (5)$$

In the frequency domain, these equations can be presented in two equivalent forms

$$Z_h(f) = Y_h(f)V(f) + \eta_h(f), \quad Z_h(f) = Y(f)V_h(f) + \eta_h(f) \quad (6)$$

where $Z_h(f)$, $Y_h(f)$, and $\eta_h(f)$ stand for DFTs of the corresponding smoothed signals. It is clear that $Y_h(f) = G_h(f)Y(f)$ and $Z_h(f) = G_h(f)Z(f)$.

We use two types of “solutions” for (6) as estimates of $y_h(x)$.

1) Regularized inverse (RI)

$$\hat{Y}_h(f) = \frac{V(-f)}{|V(f)|^2 + \varepsilon_1^2} Z_h(f) = \frac{V(-f)G_h(f)}{|V(f)|^2 + \varepsilon_1^2} Z(f). \quad (7)$$

2) Regularized Wiener inverse (RWI)

$$\hat{Y}_h(f) = \frac{V(-f)|Y(f)|^2 G_h(f)}{|V(f)Y(f)|^2 + \varepsilon_2^2 \sigma^2} Z(f). \quad (8)$$

The hat on $\hat{Y}_h(f)$ stands for the estimate of $Y(f)$. The regularization parameters $\varepsilon_1, \varepsilon_2$ control the tradeoff between the amount of a signal distortion (smoothing) and a noise suppression. If $\varepsilon_1 = \varepsilon_2 = 0$, then (7) and (8) give the naive inversion $\hat{Y}_h(f) = Z_h(f)/V(f)$. The standard Wiener inversion corresponds to $\varepsilon_2 = 1$ in (8). It yields the optimal linear mean-squared error (MSE) estimate of $Y_h(f)$ provided that the signal Y is deterministic. However, the regularization of the Wiener filter by ε_2 is an efficient tool in order to improve the filtering.

The idea behind the estimates (7) and (8) can be easier demonstrated for the pure inverse case with $\varepsilon_1 = \varepsilon_2 = 0$. Then, the estimate $\hat{y}_h(x)$ defined as the inverse Fourier transform of $\hat{Y}_h(f)$ is unbiased with respect to the smoothed $y_h(x)$, $E\{\hat{y}_h(x)\} = y_h(x)$, with the variance $\sigma_{y_h}^2(h) = \text{var}\{\hat{y}_h(x)\} = (\sigma^2)/(n_1 n_2 \|G_h(f)/V(f)\|_2^2)$, where the Euclidian norm means $\|G_h(f)/V(f)\|_2^2 = \sum_{f \in F} |G_h(f)/V(f)|^2$. The properties of the filter g_h and the PSF v are assumed to be agreed. Roughly speaking, the ratio $G_h(f)/V(f)$ and the estimate variance $\sigma_{y_h}^2$ should be finite decaying as h grows. It means that the low-pass g_h is a stronger filter than the inverse of v .

The LPA nonparametric regression with the optimal h enables a high quality approximation of y with convergence rates optimal for different classes of signals. This ability of the LPA estimates is a basic motivation behind the considered approach. However, small h are necessary to be used for high-resolution imaging of sharp details. For ill-conditioned problems it results in a large variance $\sigma_{y_h}^2$ for the pure inverse. Thus, we arrive to a necessity to apply one or another form of the regularization that works as an extra filter attenuating the noise. The estimates (7) and (8) are biased with respect to y_h if $\varepsilon_1, \varepsilon_2 > 0$. The variances of these estimates are defined by the formulas

$$\sigma_y^2(h) = \frac{\sigma^2}{n_1 n_2} \left\| \frac{V(f)G_h(f)}{|V(f)|^2 + \varepsilon_1^2} \right\|_2^2 \quad (9)$$

$$\sigma_y^2(h) = \frac{\sigma^2}{n_1 n_2} \left\| \frac{V(-f)|Y(f)|^2 G_h(f)}{|V(f)Y(f)|^2 + \varepsilon_2^2 \sigma^2} \right\|_2^2. \quad (10)$$

In our algorithm, we use the RI estimates (7) with ε_1 as small as possible leaving the main filtering to be done by the LPA with the spatially adaptive smoothing parameter h . The output of the RI is used in the RWI as a reference signal in order to estimate the power $|Y(f)|^2$ of the true image. Thus, the RI filter serves as an auxiliary pilot one to the main RWI estimator.

C. ICI Adaptive Scale Selection

The ICI rule for adaptive scale selection is derived from the accuracy analysis of the LPA. The corresponding convergence and convergence rate results have been obtained in [10] for the 1-D case. Here, we highlight basic arguments similar to those used in [10] (see also [14]–[16]), constructively leading to the ICI rule in its application to the RI and RWI algorithms. The estimation error $e(x) = y(x) - \hat{y}_h(x)$ can be represented in the form

$$|e(x)| = |y(x) - \hat{y}_h(x)| \leq |m_e(x, h)| + |e^0(x, h)| \quad (11)$$

where $\hat{y}_h(x)$ is a generic notation of the estimate given by RI or RWI algorithms, $m_e(x, h) = y(x) - E\{\hat{y}_h(x)\}$ stays for the bias and $e^0(x) = \hat{y}_h(x) - E\{\hat{y}_h(x)\}$ is a random component of the error which is Gaussian with the probability density $N(0, \sigma_y^2(h))$. It will be shown in Section III that the variance and the bias of these estimates are, respectively, decreasing and increasing functions of h . Further, if $h^*(x)$ is the ideal scale minimizing the MSE $E\{e^2(x)\}$, then

$$|m_e(x, h)| = \begin{cases} < \gamma \cdot \sigma_y(h) & \text{if } h < h^*(x) \\ > \gamma \cdot \sigma_y(h) & \text{if } h > h^*(x) \end{cases} \quad (12)$$

where γ is a constant [14], [15].

Since $|e^0(x)| \leq \chi_{1-\alpha/2} \cdot \sigma_y(h)$ holds with the probability $p = 1 - \alpha$, where $\chi_{1-\alpha/2}$ is $(1 - \alpha/2)$ th quantile of the standard Gaussian distribution, we have from (11) that with the same probability $|e(x)| \leq |m_e(x, h)| + \chi_{1-\alpha/2} \sigma_y(h)$. Using (12) for $h \leq h^*(x)$, the last inequality can be rewritten as

$$|e(x)| \leq \Gamma \cdot \sigma_y(h), \quad \Gamma = \gamma + \chi_{1-\alpha/2}. \quad (13)$$

The *confidence interval* for the biased estimate $\hat{y}_h(x)$ is introduced as follows:

$$D_h = [\hat{y}_h(x) - \Gamma \cdot \sigma_y(h), \hat{y}_h(x) + \Gamma \cdot \sigma_y(h)].$$

Then, the inequality (13) can be represented in the equivalent form $y(x) \in D_h$ provided that according to (12) $h < h^*(x)$. Consider a set H of scale parameters h

$$H = \{h_1 < h_2 < \dots < h_J\}$$

starting with a small h_1 and determining a corresponding sequence of the estimates $\hat{y}_{h_j}(x)$ and the confidence intervals $D_{h_j}, j = 1, \dots, J$.

The following is the ICI rule (statistic) used in order to obtain the adaptive scales approximating the unknown ideal $h^*(x)$ [10].

Consider the intersection of the intervals $D_{h_j}, 1 \leq j \leq i$, with increasing i , and let i^+ be the largest of those i for which the intervals $D_{h_j}, 1 \leq j \leq i$ have a point in common. This largest i^+ defines the adaptive scale h^+ and the adaptive LPA estimate as $\hat{y}^+(x) = \hat{y}_{h^+}(x), h^+(x) = h_{i^+}$.

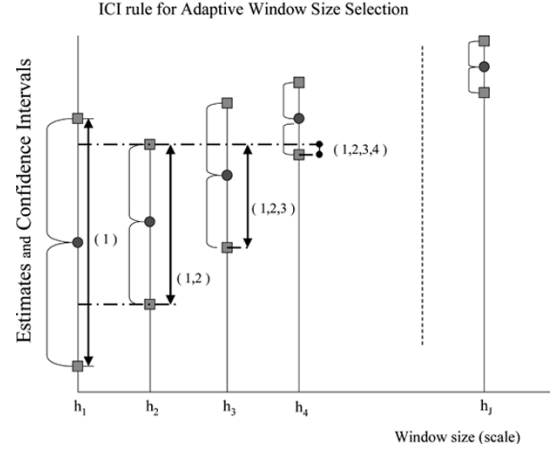


Fig. 1. Graphical illustration of the ICI rule.

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 fifth confidence interval (corresponding $h = h_5$) is empty, we obtain the adaptive window size $h^+ = h_4$.

This window size ICI selection procedure requires knowledge of the estimate and its variance only. It is equally applicable to both algorithms RI and RWI.

If the intersection of $D_{h_j}, 1 \leq j \leq i$ is not empty, it means, according to (13), that with some probability, the unknown $y(x)$ belongs to this intersection. If j is increasing, the intersection of D_{h_j} becomes smaller and possibly empty starting from a large enough h . When the intersection is empty, it is not possible to locate the unknown $y(x)$. The ICI rule defines the adaptive scale of the estimate as a maximum value of h_j when the intersection still is not empty. The theoretical analysis produced in [10] for 1-D regression and in [9] for 1-D deconvolution shows that the ICI adaptive scale estimate, provided some natural assumptions, is able to give the best possible pointwise MSE convergence rate for different classes of signals y .

D. Algorithm

1) *Basic Steps:* We develop the spatially adaptive RWI deconvolution scheme that uses the spatially adaptive RI estimate as a reference signal for the signal power estimation (see Fig. 2).

The main steps of the RI estimator can be summarized as follows.

- 1) Compute Fourier coefficients $G_h(f), Z(f), V(f)$, the RI estimate (7), and the variance (9) for $h \in H$.
- 2) Compute the estimates $\hat{y}_h(x), x \in X, h \in H$ as the inverse DFT of (7).
- 3) Apply the ICI algorithm to find the adaptive scales $h^+(x), x \in X$.
- 4) Fix the adaptive RI estimates according to the ICI rule as $\hat{y}_{RI}(x) = \hat{y}_{h^+}(x)$.

The RWI algorithm consists of the same steps as the RI algorithm with the replacement of $|Y(f)|^2$ in (8) and (10) by $|\hat{Y}_{RI}(f)|^2$.

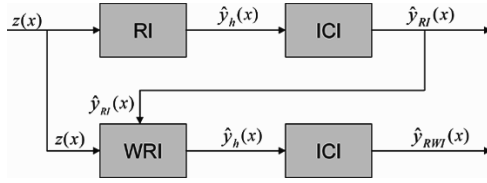


Fig. 2. LPA-ICI regularized Wiener inverse algorithm. The first line of the flowchart shows that the RI estimates $\hat{y}_h(x)$ are calculated for a set of scales $h \in H$ and ICI-algorithm is used for varying adaptive scale selection giving the RI image estimate $\hat{y}_{RI}(x)$. In the second line of the flowchart, the RWI estimates $\hat{y}_h(x)$, $h \in H$ are calculated using $\hat{y}_{RI}(x)$ as a reference signal in Wiener filtering. After that, the ICI-algorithm gives the final adaptive scale RWI estimate $\hat{y}_{RWI}(x)$.

2) *Implementation*: A symmetric window function w in (4) is a good choice if y is isotropic in a neighborhood of the estimation point. However, if y is anisotropic, as it happens near discontinuities or image edges, a nonsymmetric approximation becomes much more reasonable. To deal with the anisotropy of y , multiple nonsymmetric window LPA estimates are exploited. A neighborhood of the pixel x is separated in Q overlapping or nonoverlapping subareas. The corresponding LPA estimates $\hat{y}_h(x, q)$, $q = 1, \dots, Q$, are calculated for each of these subareas and fused together in order to yield the final estimate [15]–[17]. In this paper, we apply the directional LPA kernels which exploit different scales $h = (h_1, h_2)$ and different degrees $m = (m_1, m_2)$ of the LPA for the arguments x_1 and x_2 . The window w_h in (4) has a finite support defined as a segment of the Gaussian probability density function: $w_h(x) = f(x_1/h_1\gamma_1)f(x_2/h_2\gamma_2)$, $x_1 = 0, 1, \dots, h_1 - 1$, $x_2 = 0, 1, \dots, h_2 - 1$, where $f(x_1) = \exp(-x_1^2/2)$ and $\gamma_1, \gamma_2 > 0$. The parameters h_1 and h_2 are integer, $h_1 \geq h_2$, and define the scale and the size of the finite support. For h_2 small and $h_1 \gg h_2$ the $w_h(x)$ is a narrow support-window prolonged along the axis x_1 .

For example, for $m_1 = 2$ and $m_2 = 1$, the polynomial vector $\phi(x)$ in (4) is $\phi = (1, x_1, x_2, x_1x_2, x_1^2)$. Fig. 3 visualizes this directional kernel $g_h(x)$ for $\gamma_1 = \gamma_2 = 0.15$, $h_1 = 25$ and $h_2 = 5$. The formula (4) gives the kernel g_h provided that the number of pixels in the support is equal or larger than the size of the vector ϕ . Otherwise, it is assumed for simplicity that $m_1 = m_2 = 0$, $\phi = 1$. For $h_1 = h_2 = 1$, we always have $g_h(x) = 1$ for $x_1 = 0, x_2 = 0$ and $g_h(x) = 0$ otherwise.

We use four narrowed nonsymmetric kernels directed, respectively, to North, South, West, and East and located in the corresponding four quadrants $Q = 4$. Thus, a neighborhood of each pixel is segmented in four narrow directional subareas. The directional estimates results in much more efficient estimation as compared with the quadrant kernels exploited earlier in [15]. More about the anisotropic directional LPA estimation can be found in [16] and [17]. For each direction, the estimates $\hat{y}_h(x, q)$, $q = 1, 2, 3, 4$ are calculated independently using the ICI adaptive scale selection. There are number of ways how to aggregate these estimates into a single final one. We apply a linear weighted average defined as follows [15], [16]: $\hat{y}(x) = \sum_{q=1}^Q \lambda_q(x) \hat{y}_{h^+(x,q)}(x, q)$, $\lambda_q(x) = \text{var}(x)/\text{var}_q(x)$, $\text{var}(x) = 1/\sum_{q=1}^Q (1/\text{var}_q(x))$, where

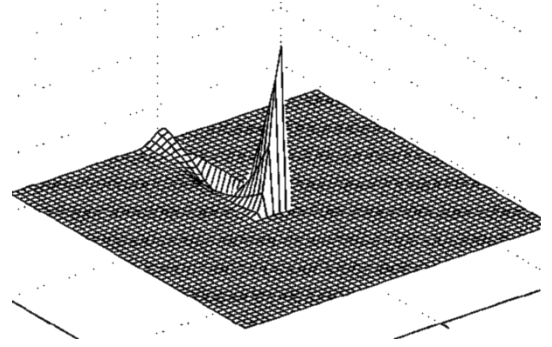


Fig. 3. Directional $g_h(x)$ kernel obtained by the LPA with the degrees $m_1 = 2, m_2 = 1$, $h_1 = 25, h_2 = 5$, $\gamma_1 = \gamma_2 = 0.15$. The window function $w(x)$ is a quadrant segment of the Gaussian distribution with $x_1, x_2 \geq 0$.

$\hat{y}_{h^+(x,q)}(x, q)$ are the kernel estimates with the ICI rule adaptive scales $h^+(x, q)$, and $\text{var}_q(x)$ are the variances of these estimates.

The variance σ^2 of the additive noise in (1) is typically unknown and must be estimated from observations. We use the median of finite differences of the observations: $\hat{\sigma}^2 = \text{median}\{|z_s - z_{s+1}|, s = 1, \dots, n_1n_2 - 1\}/(\sqrt{2} \cdot 0.6745)$, where z_s are column-wise ordered items of the observation matrix $z(x)$, $x \in X$.

3) *Complexity*: The calculation of the image estimate $\hat{y}_h(x)$ for a given h is a linear convolution requiring $N_{\text{conv}} \sim n \log n$ operations, $n = n_1n_2$. If the sectioning procedure is used for convolution, then $N_{\text{conv}} \sim n \log n_{h,J}$, where $n_{h,J}$ is a maximum size of the square mask of the kernel $g_{h,J}$. The ICI algorithm is implemented as a loop on J different scales from the set H . Its complexity is proportional to J . These calculations are repeated for each of the Q subareas of the pixel neighborhood with the fusing of these estimates to the final one. Thus, the overall complexity of calculations of the RI and RWI estimates is proportional to $J \cdot Q \cdot N_{\text{conv}}$.

III. ASYMPTOTIC THEORY

Let the object y be a function of a continuous 2-D argument and the observations (1) be sampled values of this function obtained on the grid ΔX with the sampling interval Δ on both arguments. Provided that $\Delta \rightarrow 0$, $h/\Delta \rightarrow \infty$, the discrete LPA operator (3)–(4) allows an integral representation. This representation is beneficial at least in two aspects. First, it provides a clear link with the integral smoothing operators commonly used in many applications. Second, it gives a simple and explicit dependence of the operator on h , which is useful for understanding the nature of scaled smoothing as well as for the accuracy analysis.

The convolution sum of the discrete operator can be replaced by the corresponding convolution integral of the integral operator as follows (e.g., [16])

$$\begin{aligned} y_h(x) &= (g_h \otimes y)(x) \rightarrow \int_{R^2} g(u)y(x - hu) du \\ g(x) &= w(x)\phi^T(0)\Phi^{-1}\phi(x) \\ \Phi &= \int_{R^2} w(x)\phi(x)\phi^T(x) dx. \end{aligned} \quad (14)$$

In the frequency domain, the DFT is replaced by the integral Fourier transform (IFT)

$$\begin{aligned} G(\lambda) &= \int_{R^2} g(x) \exp(-i2\pi\langle\lambda, x\rangle) dx \\ g(x) &= \int_{R^2} G(\lambda) \exp(i2\pi\langle\lambda, x\rangle) d\lambda \end{aligned} \quad (15)$$

where $\langle\lambda, x\rangle = \lambda_1 x_1 + \lambda_2 x_2$ and $\lambda = (\lambda_1, \lambda_2)$ is the 2-D continuous frequency.

The formulas for the variance are transformed to the corresponding integral forms. In particular, the variance (10) is given as follows [16]:

$$\sigma_y^2(h) = \sigma^2 \Delta^2 \int_{R^2} \frac{|V(\lambda)|^2 |Y_{\text{RI}}(\lambda)|^4 |G(h\lambda)|^2}{(|V(\lambda)Y_{\text{RI}}(\lambda)|^2 + \varepsilon_2^2 \sigma^2)^2} d\lambda. \quad (16)$$

Here, $G(\lambda) = \mathcal{F}\{g(x)\}$, $G(h\lambda) = \mathcal{F}\{g_h(x)\}$, where $\mathcal{F}\{\cdot\}$ stays for the IFT.

The estimation error is defined as a difference between the RWI estimate $\hat{y}_{\text{RWI}}(x)$ and the estimated signal $y(x)$, i.e., $e(x) = y(x) - \hat{y}_{\text{RWI}}(x)$.

Assume the following.

- 1) The object y is deterministic and belongs to a class of differentiable functions

$$C^{m+1} = \left\{ y(x): \max_{|k|=m+1} |y^{(k)}(x)| = L_{m+1}(x) \leq \bar{L}_{m+1} \right\} \quad (17)$$

where $y^{(k)}(x) = \partial_{x_1}^{k_1} \partial_{x_2}^{k_2} y(x)$, $k = (k_1, k_2)$, $|k| = k_1 + k_2$, and $L_{m+1}(x), \bar{L}_{m+1}$ are finite. The derivatives in (17) are continuous Lipschitz functions $|y^{(m+1)}(x) - y^{(m+1)}(x)| \leq L\|x - y\|^\rho$, $0 < \rho \leq 1$.

This class includes functions with bounded derivatives up to the maximum order $m+1$. Polynomial of the power m belong to this class. Indeed, if $\bar{L}_{m+1} = 0$, then for all x and $|k| = m+1$, the derivatives $y^{(k)}(x) = 0$ and $y(x)$ is a polynomial of the power m .

- 2) The discrete convolution kernel in (1) is a sampled continuous function $v(x)$, with $V(\lambda) = \mathcal{F}\{v(x)\}$. It is assumed that $V(\lambda)$ is polynomially decaying (ordinary smooth PSF [9]). There exist real positive $\bar{c}_0, \underline{c}_0, c_1, A, \alpha$ such that

$$\begin{aligned} \bar{c}_0 &\geq |V(\lambda)| \|\lambda\|^\alpha \geq \underline{c}_0, \quad \forall \|\lambda\| > A, \min_{\|\lambda\| \leq A} |V(\lambda)| \\ &= c_1, \alpha > 0. \end{aligned} \quad (18)$$

- 3) The estimator kernel g defined by (4) has IFT $G(\lambda)$ and the following conditions hold:

$$\begin{aligned} B_{Y,\alpha}^2 &= \int_{R^2} |Y| \cdot \|\lambda\|^{2\alpha} d\lambda < \infty \\ B_{G,\alpha}^2 &= \int_{R^2} |G| \cdot \|\lambda\|^{2\alpha} d\lambda < \infty \\ B_{V,Y}^2 &= \int_{R^2} |V|^2 |Y|^4 d\lambda < \infty. \end{aligned} \quad (19)$$

The $L_{m+1}(x)$ in 1) is an upper bound for the derivatives of order $m+1$; 2) defines a decaying rate of $|V|$ as $\|\lambda\| \rightarrow \infty$. It is possible to restore the true image provided that its spectrum $|Y|$ is not wider than the pass-band of the PSF. Details of the image “killed” by the PSF cannot be restored. The first inequality in (19) formalizes this requirement. It suffices $\int_{R^2} |Y|/|V|^2 d\lambda < \infty$. In a similar way, the second inequality in (19) suffices $\int_{R^2} |G|/|V|^2 d\lambda < \infty$.

- 4) The object function is also polynomial decaying with the conditions $\bar{d}_0 \geq |Y(\lambda)| \|\lambda\|^\beta \geq \underline{d}_0, \forall \|\lambda\| > A, \min_{\|\lambda\| \leq A} |Y(\lambda)| = d_1$. It is a technical (not principal) assumption allowing to simplify the analysis.

The following proposition shows the orders of the estimation bias and variance with respect to the scale h of the RI-RWI algorithm.

Proposition 1: Consider the RI-RWI estimate. Assume that 1)–4) hold. Then, for the asymptotics with small $\Delta, h, \Delta/h, \Delta/h^{\alpha+1}, \sigma, \varepsilon_1^2, \varepsilon_2^2$, the following upper bounds are valid.

- 1) The estimation bias

$$|E\{e(x)\}| \leq |\tilde{m}_e(x, h)| + o(h^{m+1}) + o(\varepsilon_2^2) \quad (20)$$

$$|\tilde{m}_e(x, h)| = h^{m+1} L_{m+1}(x) a_g,$$

$$a_g = \sum_{|k|=m+1} \frac{1}{k!} \left| \int_{R^2} g(u) u^k du \right|. \quad (21)$$

- 2) The estimation variance

$$\text{var}\{e(x)\} \leq \tilde{\sigma}_y^2(h) + o\left(\Delta^2 h^{-2(\alpha+1)}\right)$$

$$\tilde{\sigma}_y^2(h) = d_p \sigma^2 \Delta^2 h^{-2(\alpha+1)}$$

$$d_p = \frac{\bar{c}_0^2 \bar{d}_0^4}{(\underline{c}_0^2 \underline{d}_0^2 + \sigma^2 \varepsilon_2^2 A^{2\alpha+2\beta})^2} B_{G,\alpha}^2. \quad (22)$$

The proof of the proposition is outlined in the Appendix.

It follows from (20)–(22) that the main term of the MSE risk allows the following upper bound:

$$E\{e^2(x, h)\} \leq \bar{r}^2(x, h), \quad \bar{r}^2(x, h) = (\tilde{m}_e(x, h))^2 + \tilde{\sigma}_y^2(h). \quad (23)$$

The bias and variance are, respectively, increasing and decreasing functions of h . The quadratic risk is convex on h and achieves its minimal value at

$$\begin{aligned} h^*(x) &= \left(\frac{d_p \sigma^2 \Delta^2}{L_{m+1}^2(x) a_g^2} \gamma^2 \right)^{1/(2m+2\alpha+4)} \\ \gamma &= \sqrt{\frac{\alpha+1}{m+1}}. \end{aligned} \quad (24)$$

It can be derived from (20)–(22) and (24) that $|\tilde{m}_e(x, h^*)| = \gamma \tilde{\sigma}_y(h^*)$ for $h = h^*(x)$ and

$$|\tilde{m}_e(x, h)| = \begin{cases} < \gamma \cdot \tilde{\sigma}_y(h) & \text{if } h < h^*(x) \\ > \gamma \cdot \tilde{\sigma}_y(h) & \text{if } h > h^*(x) \end{cases}. \quad (25)$$

The parameter γ is a ratio of the bias $|\tilde{m}_e|$ to the standard deviation $\tilde{\sigma}_y$ when both of them are calculated for the ideal scale $h^*(x)$. This ratio is a constant depending only on the order m of the LPA and the smoothness of the PSF given by α .

The ideal $h^*(x)$ depends on the $|m+1|$ th derivatives of $y(x)$ (through $L_{m+1}(x)$) and, in this way, the ideal scale appeared to be spatially varying. It proves (12).

The test of the hypothesis $h \leq h^*(x)$ is a basic idea of the ICI rule. It is used in order to obtain the adaptive values of h close to the ideal $h^*(x)$.

Inserting the ideal scale parameter $h^*(x)$ (24) into (23), we derive the order of the MSE risk with respect to the sampling interval Δ : $r(x, h^*(x)) = O(\Delta^{(2(m+1))/(m+\alpha+2)})$. Let the sampling interval Δ is inversely proportional to the number of samples. Then, for $n = n_1 n_2 \sim 1/\Delta^2$, we have

$$r(x, h^*(x)) = O\left(n^{-\frac{m+1}{m+\alpha+2}}\right). \quad (26)$$

This last formula shows the ideal MSE convergence rate with respect to the number of samples in the image. A larger m improves the convergence rate. The parameter α in (26) shows the influence of the PSF. The convergence rate becomes slower for larger values of α , i.e., for a smoother PSF. With $\alpha = 0$, the formula (26) coincides with the well-known results for the best convergence rate of the nonparametric regression estimation in direct observations which do not require deblurring [6], [13], [15], [16].

The proposition results are obtained provided that $V(\lambda)$ is not vanishing, $V(\lambda) \neq 0$ [assumption (18)]. For convergence of the algorithm, the regularization parameters should approach zero value sufficiently quickly. It can be shown for the RI-RWI algorithm that these parameters should be of the orders $\varepsilon_1^2 = O(n^{-(m+1)/((\alpha+m+2)2)})$ and $\varepsilon_2^2 = O(n^{-(\alpha+m+2+\beta/2)/((\alpha+m+2)2)})$. The formula (26) gives the oracle convergence rate as it is assumed that $y(x)$ or $L_{m+1}(x)$ are known and used for the selection of the ideal $h^*(x)$. It follows from some general results of the adaptive estimation theory (e.g., [10]) that if the estimated function is unknown and the adaptive scale algorithm is used, the convergence rate cannot be better than $r(x, h^*(x)) = O((\ln n/n)^{(m+1)/(m+\alpha+2)})$. The factor $\ln n$ in this formula is a “price” of the adaptivity. The quality of the adaptive estimate coincides up to this log factor with the convergence rate of the oracle estimate using the estimated function values for the scale selection.

IV. EXPERIMENTS

In this section, we present simulation results demonstrating the performance of the developed algorithm in comparison with some recent advanced wavelet-based methods [7], [11], [21]. The term “state-of-art” has been used in [7] for these methods demonstrating high values of imaging performance criteria. The LPA is applied with the following parameters: the directional LPA kernels as described in Section II-D.2 of the degrees $m_1 = 1, m_2 = 0$ for the RI algorithm and $m_1 = 0, m_2 = 0$ for the RWI algorithm. The scale $h_2 = 1$ is fixed and h_1 is adaptive varying $h_1 \in H = \{1, 3, 5, 8, 13\}$. The parameters γ_1, γ_2 of the window $w_h(x)$ are different for different scales: $\gamma_1 = \gamma_2 = 0.16h_1$. It allows to use the wider kernels for larger values of h_1 .

TABLE I
ISNR FOR FOUR EXPERIMENTS

Method	Experiment	1	2	3	4
<i>RI - RWI</i>		7.84	7.31	5.54	3.76
<i>EM</i> (Figueiredo and Nowak) [7]		7.59	6.93	4.88	2.94
<i>ForWaRD</i> (Neelamani et al.) [21]		7.40	6.75	5.07	2.98

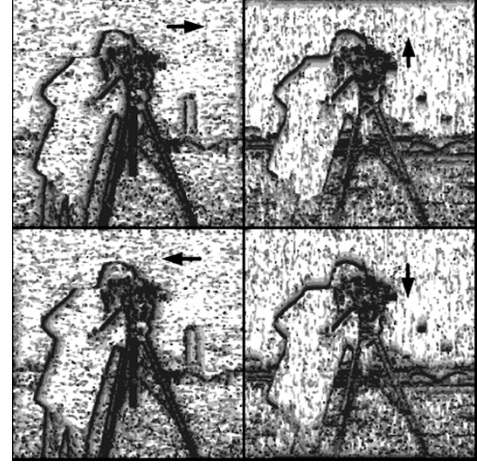


Fig. 4. ICI adaptive scales h^+ for four directions shown by arrows.

The design parameters of the LPA-ICI deblurring are the ICI thresholds Γ_1, Γ_2 and the regularization parameters $\varepsilon_1, \varepsilon_2$, respectively, for RI (7) and RWI (8) algorithms. In multiple experiments with different images, we found the following values of these parameters which overall enable a good performance: $\Gamma_1 = \Gamma_2 = 1.0, \varepsilon_1 = 0.014, \varepsilon_2 = 0.35$.

Further improvement can be achieved by a proper tuning these parameters. For adaptive tuning, a loss function can be used in the form proposed in [21] for the adaptive regularization parameter selection.

The following standard criteria are used: root mean-squared error (RMSE): $\text{RMSE} = \sqrt{(1/\#) \sum_x (y(x) - \hat{y}(x))^2}$; SNR in decibels: $\text{SNR} = 10 \log_{10} \sum_x |y(x)|^2 / \sum_x |y(x) - \hat{y}(x)|^2$; improvement in SNR (ISNR) in decibels: $\text{ISNR} = 20 \log_{10}(\hat{\sigma}/\text{RMSE})$; peak signal-to-noise ratio (PSNR) in decibels: $\text{PSNR} = 20 \log_{10}(\max_x |y(x)|/\text{RMSE})$; mean absolute error (MAE): $\text{MAE} = (1/\#) \sum_x |y(x) - \hat{y}(x)|$; maximum absolute error: $\text{MAXDIF} = \max_x |y(x) - \hat{y}(x)|$. Most of these criteria actually define the accuracy of approximation of the image intensity function. There is no one-to-one link between the image quality and the above criteria. It was noticed that different criteria sometimes show quite different optimal values for the design parameters. Thus, a visual inspection, which, of course, is quite subjective, continues to be the most important final performance criterion.

In the first set of experiments, we consider the setup of [21]. The test signal is the 256×256 “Cameraman” image (8-bit gray-scale) corrupted by an additive zero-mean Gaussian noise. The blurred SNR (BSNR) is defined in decibels as $10 \log_{10}[\sum_x ((y \circledast v)(x) - (1/(n_1 n_2)) \sum_x (y \circledast v)(x))^2 / \sigma^2 n_1 n_2]$ with $n_1 = n_2 = 256$, BSNR = 40 dB. The discrete-space blur convolution PSF is a uniform 9×9 box car. Table I shows ISNR for these experiments in column Experiment 1. The best figure 7.59 dB up until now was obtained in

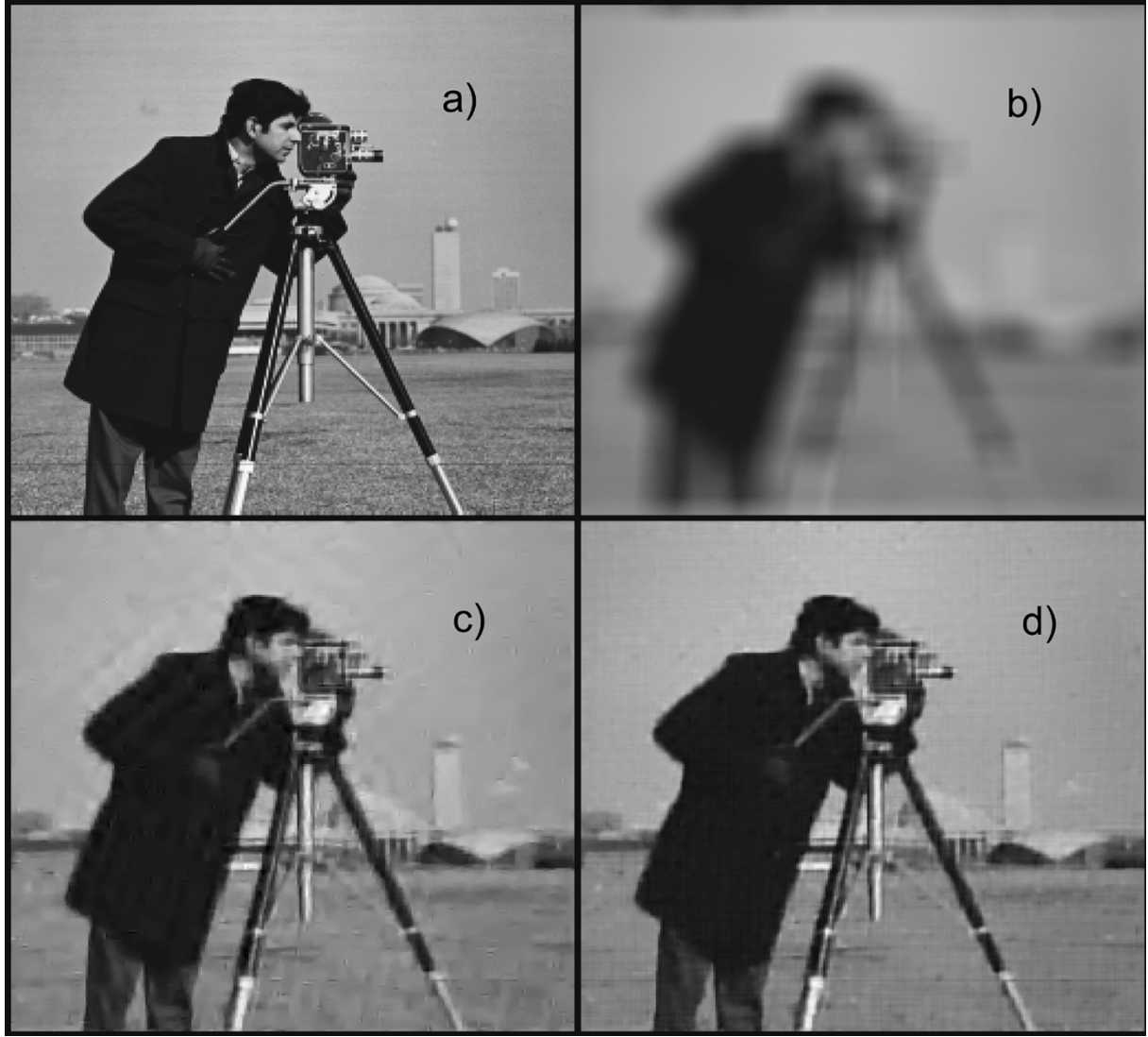


Fig. 5. Reconstruction of "Cameraman" image. (a) True image. (b) Noisy blurred image. PSF is a 19×19 uniform box-car, BSNR = 40 dB. (c) *ForWaRD* algorithm ISNR = 6.5 dB. (d) LPA-ICI algorithm ISNR = 7.0 dB. Visually, the comparison is in favor of LPA-ICI algorithm.

[7]. Our RWI algorithm yields a better value 7.84 which is also better than 7.4 dB obtained by the *ForWaRD* algorithm.

The Matlab implementation of the *ForWaRD* algorithm is available at www.rice.edu/software. Comparative experiments with *ForWaRD* and our algorithm have been done for number of standard test-images: Lena, Theater, Boats, Birthday, etc. We compare the algorithms using all above criteria. Overall, in criteria values, the RWI algorithm consistently demonstrates a better performance than *ForWaRD*. Concerning the visual inspection RWI significantly outperforms *ForWaRD*, providing reconstructed images without artifacts, flaws, etc., common for wavelet based thresholding using image segmentation. As an illustration, Fig. 5 presents a reconstruction of "Cameraman" produced by the RWI and *ForWaRD* algorithms with quite a clear advantage of the RWI algorithm. The corresponding adaptive window sizes (scales) $h^+(x)$ for four directions indicated by arrows are demonstrated in Fig. 4. Small and large windows are shown by black and white, respectively. The obtained window sizes actually correspond to the intuitively clear behavior of the varying window size relevant to the smoothing if the true image

is known. Thus, the adaptive window sizes delineate the true image 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 behavior of the estimates. Isolated black points in Fig. 4 correspond to random small window sizes erroneously given by the ICI rule. These isolated spikes have different locations into four different windows and do not influence the final image reconstruction shown in Fig. 5.

We have performed experiments on some artifact images. In particular, for the "Box" image (binary image 64×64 , black background and white central square 32×32 , BSNR = 40 dB, 9×9 box-car PSF) we yield ISNR = 17.2 dB for RWI versus ISNR = 14.1 dB for *ForWaRD*.

In the second set of tests, we replicate the experiential conditions of [7], the test-image "Cameraman," the PSF: $v(k_1, k_2) = (1 + k_1^2 + k_2^2)^{-1}$, $k_1, k_2 = -7, \dots, 7$, and the noise variances are $\sigma^2 = 2$ (BSNR = 32 dB) and $\sigma^2 = 8$ (BSNR = 26 dB). The results for ISNR are shown in Table I in columns Experiment 2 and Experiment 3, respectively. The RWI algorithm demonstrates the best values of ISNR.

In the last set of experiments, we use the simulation conditions from [7]: the 512×512 test-image “Lena,” $v(k_1, k_2)$ is a 5×5 separable filter with the weights $[1, 4, 6, 4, 1]/16$ in both horizontal and vertical directions; white Gaussian additive noise with $\sigma = 7$ (BSNR = 15.93 dB). The results for ISNR are shown in column Experiment 4 of Table I. The RWI shows a significant improvement in comparison with the competitive algorithms. Overall, the simulation results shows that the new developed RWI algorithm demonstrates a good performance and outperforms some of the best in the field algorithms.

V. CONCLUSION

The new adaptive scale deblurring technique is developed. It is based on the directional LPA applied for design of the low-pass filters joined with the regularized inversion and Wiener inversion. The ICI rule gives the varying adaptive scale selection which transforms the originally linear LPA and inversion filters to nonlinear adaptive systems. The simulation confirms a state-of-art performance of the proposed algorithm.

The LPA allows a number of interesting further developments. In particular, an optimization of the kernel filters $g_h(x)$ can be produced in order to improve the performance of the algorithms. Further, the LPA can be used in order to design the differentiation filters [16]. Then, replacing the estimation kernel g_h by the differentiation kernel the above estimates can be transformed in adaptive scale differentiation algorithms with efficient estimation of the derivatives of the image intensity. These algorithms can be used for edge detection, image improvements, recognition problems, etc.

APPENDIX

Proof of Proposition: In the frequency domain, the estimation error of the RWI algorithm is of the form

$$e(\lambda) = \left(1 - \frac{G(\lambda h)|\hat{Y}_{\text{RI}}(\lambda)|^2|V(\lambda)|^2}{|\hat{Y}_{\text{RI}}(\lambda)|^2|V(\lambda)|^2 + \sigma^2\varepsilon_2^2}\right)Y(\lambda) - \frac{G(\lambda h)|\hat{Y}_{\text{RI}}(\lambda)|^2V(-\lambda)}{|\hat{Y}_{\text{RI}}(\lambda)|^2|V(\lambda)|^2 + \sigma^2\varepsilon_2^2}\eta(\lambda). \quad (27)$$

Represent the RI estimate $\hat{Y}_{\text{RI}}(\lambda)$ as $\hat{Y}_{\text{RI}}(\lambda) = Y(\lambda) + e_{\text{RI}}(\lambda)$, where $e_{\text{RI}}(\lambda)$ is a small estimation error. Then, $|\hat{Y}_{\text{RI}}(\lambda)|^2 \simeq |Y(\lambda)|^2 + 2\text{Re}(e_{\text{RI}}(\lambda)Y(-\lambda))$ and substituting it in (27), we can calculate $e(\lambda)$ using two terms of the Taylor series

$$e \simeq e|_0 + [\partial e / \partial |\hat{Y}_{\text{RI}}|^2]|_0 \cdot 2\text{Re}(e_{\text{RI}}(\lambda)Y(-\lambda)). \quad (28)$$

Here, $|_0$ means that $e_{\text{RI}}(\lambda) = 0$ in the corresponding items of the expression. It can be seen that $[\partial e / \partial |\hat{Y}_{\text{RI}}|^2]|_0 = 0(\varepsilon_2^2)$. Multiplied by $|e_{\text{RI}}(\lambda)|$, it gives the term smaller in order than $e|_0$. Then, we can evaluate the MSE risk using only the first term in (28) assuming that the power of the reference signal $|\hat{Y}_{\text{RI}}(\lambda)|^2$ in (27) is equal to the accurate value $|Y(\lambda)|^2$.

It follows from (27) that

$$\text{var}\{e(x)\} \simeq \sigma^2\Delta^2 \int_{\mathbb{R}^2} \left| \frac{G(\lambda h)|Y(\lambda)|^2V(-\lambda)}{|Y(\lambda)|^2|V(\lambda)|^2 + \sigma^2\varepsilon_2^2} \right|^2 d\lambda \quad (29)$$

$$E\{e(x)\} \simeq \mathcal{F}^{-1}\{(Y(\lambda) - G(\lambda h)Y(\lambda)) + \frac{\sigma^2\varepsilon_2^2}{|V(\lambda)|^2|Y(\lambda)|^2 + \sigma^2\varepsilon_2^2}G(\lambda h)Y(\lambda)\}. \quad (30)$$

Consider the variance (29) in the following analysis:

$$\begin{aligned} & \int_{\mathbb{R}^2} \frac{|G(h\lambda)|^2|V(\lambda)|^2|Y(\lambda)|^4}{(|V(\lambda)|^2|Y(\lambda)|^2 + \sigma^2\varepsilon_2^2)^2} d\lambda \\ &= \int_{\|\lambda\| \leq A} \frac{|G(h\lambda)|^2|V(\lambda)|^2|Y(\lambda)|^4}{(|V(\lambda)|^2|Y(\lambda)|^2 + \sigma^2\varepsilon_2^2)^2} d\lambda \\ &+ \int_{\|\lambda\| > A} \frac{|G(h\lambda)|^2|V(\lambda)|^2|Y(\lambda)|^4}{(|V(\lambda)|^2|Y(\lambda)|^2 + \sigma^2\varepsilon_2^2)^2} d\lambda. \end{aligned} \quad (31)$$

Using assumptions 2) and 4) and $|G(\lambda h)| = 1 + 0(h)$ for small A , we obtain

$$\begin{aligned} & \int_{\|\lambda\| \leq A} \frac{|G(h\lambda)|^2|V(\lambda)|^2|Y(\lambda)|^4}{(|V(\lambda)|^2|Y(\lambda)|^2 + \sigma^2\varepsilon_2^2)^2} d\lambda \\ & \leq \int_{\|\lambda\| \leq A} \frac{|G(\lambda h)|^2|V(\lambda)|^2|Y(\lambda)|^4}{(c_1^2d_1^2 + \sigma^2\varepsilon_2^2)^2} d\lambda \\ & \leq \int_{\|\lambda\| \leq A} \frac{|V(\lambda)|^2|Y(\lambda)|^4}{(c_1^2d_1^2 + \sigma^2\varepsilon_2^2)^2} d\lambda \leq \frac{B_{V,Y}^2}{(c_1^2d_1^2 + \sigma^2\varepsilon_2^2)^2} \end{aligned} \quad (32)$$

and

$$\begin{aligned} & \int_{\|\lambda\| > A} \frac{|G(h\lambda)|^2|V(\lambda)|^2|Y(\lambda)|^4}{(|V(\lambda)|^2|Y(\lambda)|^2 + \sigma^2\varepsilon_2^2)^2} d\lambda \\ & \leq \int_{\|\lambda\| > A} \frac{|G(h\lambda)|^2\tilde{c}_0^4\tilde{d}_0^4}{(\tilde{c}_0^2\tilde{d}_0^2 + \sigma^2\varepsilon_2^2\|\lambda\|^{2\alpha+2\beta})^2} \|\lambda\|^{2\alpha} d\lambda \\ & \leq \frac{\tilde{c}_0^4\tilde{d}_0^4h^{-(2\alpha+2)}}{(\tilde{c}_0^2\tilde{d}_0^2 + \sigma^2\varepsilon_2^2A^{2\alpha+2\beta})^2}. \end{aligned} \quad (33)$$

Inserting (32) and (33) into (31), we obtain (22).

Consider the bias $E\{e(x)\}$. The first summand in (30) has a form $\mathcal{F}^{-1}\{Y(\lambda) - G(\lambda h)Y(\lambda)\} = y(x) - \int_{\mathbb{R}^2} g(u)y(x - hu)du$. The standard technique based on the Taylor series for $y(x - hu)$ with respect to h gives the first summand in (20) (see [15] and [16]). It can be verified that the second summand in (30) is small of the order $0(\varepsilon_2^2)$. It gives (20). \square

ACKNOWLEDGMENT

The authors would like to thank the three anonymous referees for helpful and stimulating comments. They would also like to thank A. Foi for his work in algorithm implementation and simulation experiments.

REFERENCES

- [1] F. Abramovich and B. Silverman, “Wavelet decomposition approaches to statistical inverse problems,” *Biometrika*, vol. 85, pp. 115–129, 1998.
- [2] E. J. Candès and D. L. Donoho, “Recovering edges in ill-posed inverse problems: Optimality of curvelet frames,” *Ann. Stat.*, vol. 30, pp. 784–842, 2000.

- [3] L. Cavalier and A. Tsybakov, "Sharp adaptation for inverse problems with random noise," *Probl. Theory Related Fields*, vol. 123, pp. 323–354, 2002.
- [4] D. L. Donoho, "Non-linear solution of linear inverse problems by wavelet-vaguelette decomposition," *Appl. Comput. Harm. Anal.*, vol. 2, pp. 101–126, 1995.
- [5] —, "Wedgelets: Nearly minimax estimation of edges," *Ann. Stat.*, vol. 27, pp. 859–897, 1999.
- [6] J. Fan and I. Gijbels, *Local Polynomial Modeling and Its Application*. London, U.K.: Chapman & Hall, 1996.
- [7] M. A. T. Figueiredo and R. D. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Trans. Image Process.*, vol. 12, no. 8, pp. 906–916, Aug. 2003.
- [8] L. Ganesan and P. Bhattacharyya, "Edge detection in untextured and textured images—A common computational framework," *IEEE Trans. Syst., Man, Cybern. B: Cybern.*, vol. 27, no. 2, pp. 823–834, May 1997.
- [9] A. Goldenshluger, "On pointwise adaptive nonparametric deconvolution," *Bernoulli*, vol. 5, pp. 907–925, 1999.
- [10] A. Goldenshluger and A. Nemirovski, "On spatial adaptive estimation of nonparametric regression," *Math. Meth. Stat.*, vol. 6, pp. 135–170, 1997.
- [11] A. Jalobeanu, N. Kingsbury, and J. Zerubia, "Image deconvolution using hidden Markov tree modeling of complex wavelet packets," presented at the IEEE Int. Conf. Image Processing, Thessaloniki, Greece, 2001.
- [12] J. Kalifa, S. Mallat, and B. Rouge, "Deconvolution by thresholding in mirror wavelet bases," *IEEE Trans. Image Process.*, vol. 12, no. 4, pp. 446–457, 2003.
- [13] V. Katkovnik, *Nonparametric Identification and Smoothing of Data (Local Approximation Methods)*. Moscow, Russia: Nauka, 1985.
- [14] —, "A new method for varying adaptive bandwidth selection," *IEEE Trans. Signal Process.*, vol. 47, no. 9, pp. 2567–2571, Sep. 1999.
- [15] V. Katkovnik, K. Egiazarian, and J. Astola, "Adaptive window size image de-noising based on intersection of confidence intervals (ICI) rule," *J. Math. Imag. Vis.*, vol. 16, no. 3, pp. 223–235, 2002.
- [16] —, *Adaptive Varying Scale Methods in Image Processing*. Tampere, Finland: Tampere Int. Center Signal Processing, 2003.
- [17] V. Katkovnik, A. Foi, K. Egiazarian, and J. Astola, "Directional varying scale approximations for anisotropic signal processing," presented at the EUSIPCO, 2004.
- [18] O. Lepski, E. Mammen, and V. Spokoiny, "Ideal spatial adaptation to inhomogeneous smoothness: An approach based on kernel estimates with variable bandwidth selection," *Ann. Stat.*, vol. 25, no. 3, pp. 929–947, 1997.
- [19] C. Loader, *Local Regression, and Likelihood*. New York: Springer, 1999.
- [20] S. Mallat, *A Wavelet Tour of Signal Processing*. New York: Academic, 1999.
- [21] R. Neelamani, H. Choi, and R. G. Baraniuk, "Forward: Fourier-wavelet regularized deconvolution for ill-conditioned systems," *IEEE Trans. Image Process.*, vol. 52, no. 2, pp. 418–433, Feb. 2004.
- [22] J. Polzehl and V. Spokoiny, "Image denoising: Pointwise adaptive approach," *Ann. Stat.*, vol. 31, no. 1, 2003.
- [23] A. N. Tikhonov and V. Y. Arsenin, *Solution of Ill-Posed Problems*. New York: Wiley, 1977.



Vladimir Katkovnik received the M.Sc., Ph.D., and D.Sc. degrees in technical cybernetics from the Leningrad Polytechnic Institute, Leningrad, Russia, in 1960, 1964, and 1974, respectively.

From 1964 to 1991, he held the positions of Associate Professor and Professor at the Department of Mechanics and Control Processes, Leningrad Polytechnic Institute. From 1991 to 1999, he was a Professor of statistics with the Department of the University of South Africa, Pretoria. From 2001 to 2003, he was a Professor of mechatronics with the Kwangju Institute of Science and Technology, Korea. From 2000 to 2001, he was a Visiting Professor with the Signal Processing Laboratory, Tampere University of Technology, Tampere, Finland. He has published seven books and more than 150 papers. His research interests include stochastic signal processing, linear and nonlinear filtering, nonparametric estimation, imaging, nonstationary systems, and time-frequency analysis.



Karen Egiazarian (SM'96) was born in Yerevan, Armenia, in 1959. He received the M.Sc. degree in mathematics from Yerevan State University in 1981, the Ph.D. degree in physics and mathematics from Moscow State University, Moscow, Russia, in 1986, and the D.Tech. degree from the Tampere University of Technology (TUT), Tampere, Finland, in 1994.

He has been Senior Researcher with the Department of Digital Signal Processing, Institute of Information Problems and Automation, National Academy of Sciences of Armenia. Since 1996, he has been an Assistant Professor with the Institute of Signal Processing, TUT, where he is currently a Professor, leading the Spectral and Algebraic Methods in DSP group. His research interests are in the areas of applied mathematics, signal processing, and digital logic.



Jaakko Astola (F'00) received the B.Sc., M.Sc., Licentiate, and Ph.D. degrees in mathematics (specializing in error-correcting codes) from Turku University, Turku, Finland, in 1972, 1973, 1975, and 1978, respectively.

From 1976 to 1977, he was with the Research Institute for Mathematical Sciences of Kyoto University, Kyoto, Japan. Between 1979 and 1987, he was with the Department of Information Technology, Lappeenranta University of Technology, Lappeenranta, Finland, holding various teaching positions in mathematics, applied mathematics, and computer science. In 1984, he was a Visiting Scientist at Eindhoven University of Technology, Eindhoven, The Netherlands. From 1987 to 1992, he was an Associate Professor in applied mathematics at Tampere University, Tampere, Finland. Since 1993, he has been a Professor of signal processing and Director of the Tampere International Center for Signal Processing, leading a group of about 60 scientists. His research interests include signal processing, coding theory, spectral techniques and statistics.

Dr. Astola was nominated as Academy Professor by the Academy of Finland (2001 to 2006).