# Pointwise Shape-Adaptive DCT for High-Quality Denoising and Deblocking of Grayscale and Color Images

Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian, *Senior Member, IEEE*

*Abstract*—The shape-adaptive DCT (SA-DCT) transform can be computed on a support of arbitrary shape, but retains a computational complexity comparable to that of the usual separable block-DCT (B-DCT). Despite the near-optimal decorrelation and energy compaction properties, application of the SA-DCT has been rather limited, targeted nearly exclusively to video compression.

In this paper we present a novel approach to image filtering based on the SA-DCT. We use the SA-DCT in conjunction with the Anisotropic Local Polynomial Approximation (LPA) - Intersection of Confidence Intervals (ICI) technique, which defines the shape of the transform's support in a pointwise adaptive manner. The thresholded or attenuated SA-DCT coefficients are used to reconstruct a local estimate of the signal within the adaptive-shape support. Since supports corresponding to different points are in general overlapping, the local estimates are averaged together using adaptive weights that depend on the region's statistics.

This approach can be used for various image processing tasks. In this paper we consider in particular image denoising and image deblocking and deringing from block-DCT compression. A special structural constraint in luminance-chrominance space is also proposed to enable an accurate filtering of color images. Simulation experiments show a state-of-the-art quality of the final estimate, both in terms of objective criteria and visual appearance. Thanks to the adaptive support, reconstructed edges are clean, and no unpleasant ringing artifacts are introduced by the fitted transform.

*Index Terms*—shape-adaptive, DCT, denoising, deblocking, deringing, anisotropic.

## I. INTRODUCTION

The two-dimensional separable block-DCT (B-DCT), computed on a square or rectangular support, is a well established and very efficient transform in order to achieve a sparse representation of image blocks. For natural images, its decorrelating performance is close to that of the optimum Karhunen-Loève transform. Thus, the B-DCT has been successfully used as the key element in many compression and denoising applications. However, in presence of singularities or edges such near-optimality fails. Because of the lack of sparsity, edges cannot be coded or restored effectively, and ringing artifacts arising from the Gibbs phenomenon become visible.

In the last decade, significant research has been made towards the development of region-oriented, or *shape-adaptive*, transforms. The main intention is to construct a system (frame, basis, etc.) that can efficiently be used for the analysis and synthesis of arbitrarily shaped image segments, where the data exhibit some uniform behavior.

Initially, Gilge [19], [20] considered the orthonormalization of a (fixed) set of generators restricted to the arbitrarily shaped region of interest. These generators could be a basis of polynomials or – for example – a B-DCT basis, thus yielding a "shape-adapted" DCT transform. Orthonormalization can be performed by the standard Gram-Schmidt procedure and the obtained orthonormal basis is supported on the region. Because the region-adapted basis needs to be recalculated for each differently shaped region and because the basis elements are typically non-separable, the overall method presents a rather high computational cost. While even today it is considered as one of the best solutions to the region-oriented transforms problem, Gilge's approach is clearly unsuitable for real-time applications, and faster transforms were sought.

A more computationally attractive approach, namely the *shape-adaptive DCT* (SA-DCT), has been proposed by Sikora et al. [47], [49]. The SA-DCT is computed by cascaded application of one-dimensional varying-length DCT transforms first on the columns and then on the rows that constitute the considered region, as shown in Figure 1. Thus, the SA-DCT does not require costly matrix inversions or iterative orthogonalizations and can be interpreted as a direct generalization of the classical 2D B-DCT transform. In particular, the SA-DCT and the B-DCT (which is separable) have the same computational complexity and in the special case of a square the two transforms exactly coincide. Therefore, the SA-DCT has received considerable interest from the MPEG community, eventually becoming part of the MPEG-4 standard [32], [36]. The recent availability of low-power SA-DCT hardware platforms (e.g. [5],[30],[31]) makes this transform an appealing choice for many image- and video-processing tasks.

The SA-DCT has been shown [47], [48], [4], [27] to provide a compression efficiency comparable to those of more computationally complex transforms, such as [20]. The good decorrelation and energy compaction properties on which this efficiency depends are also the primary characteristics sought for any transform-domain denoising algorithm. In this sense, the SA-DCT features a remarkable potential not only for video compression applications, but also for image and video

denoising.

However, this potential has been apparently ignored by the image denoising and restoration community. While this indifference may seem rather surprising, there are sound reasons that can justify it.

The use of a transform with a shape-adaptive support involves actually two separate problems: not only the transform should adapt to the shape (i.e. a shape-adaptive transform), but the shape itself must adapt to the image features (i.e. an image-adaptive shape). The first problem has found a very satisfactory solution in the SA-DCT transform. How to deal with the second problem depends on the considered application.

The shape-adaptive coding of *noise-free* video objects always assumes that the boundary of these objects is known. This information can be obtained either from a-priori knowledge (e.g. motion estimation, chroma keying, layered structure), or it can be estimated from the data with one of the many automated segmentation algorithms suitable for this purpose (e.g. [37]). On the contrary, obtaining an accurate and robust segmentation of noisy data constitutes an extremely more complex task than the region-oriented coding itself. Unlike in video coding, such a segmentation cannot be reasonably assumed to be known a-priori. It must be noted that conventional segmentation (or local-segmentation) techniques which are employed for video processing are not suitable for degraded (noisy, blurred, highly compressed, etc.) data. This very aspect may be identified as the principal reason why the SA-DCT had not been used for the restoration of noisy images.

In our approach, we use the SA-DCT in conjunction with the Anisotropic Local Polynomial Approximation (LPA) - Intersection of Confidence Intervals (ICI) [26], [24], [21], [18], a technique purposely designed to work accurately with noisy data. By comparing varying-scale directional kernel estimates, this technique adaptively selects, for each point in the image, a set of directional adaptive-scales. The length of the support (i.e. the window size) of the corresponding adaptive-scale kernels define the shape of the transform's support in a pointwise-adaptive manner. Examples of such neighborhoods are shown in Figures 2, 4, 5, and 13.

For each one of these neighborhoods a SA-DCT is performed. The hard-thresholded SA-DCT coefficients are used to reconstruct a local estimate of the signal within the adaptive-shape support. By using the adaptive neighborhoods as support for the SA-DCT, we ensure that data are represented sparsely in the transform domain, allowing to effectively separate signal from noise using hard-thresholding.

Since supports corresponding to different points are in general overlapping (and thus generate an overcomplete representation of the signal), the local estimates are averaged together using adaptive weights that depend on the local estimates' statistics. In this way we obtain an adaptive estimate of the whole image.

Once this global estimate is produced, it can be used as reference estimate for an empirical Wiener filter in SA-DCT domain. Following the same adaptive averaging procedure as for hard-thresholding, we arrive to the final Anisotropic LPA-ICI-driven SA-DCT estimate. We term our approach "*Pointwise SA-DCT filtering*".
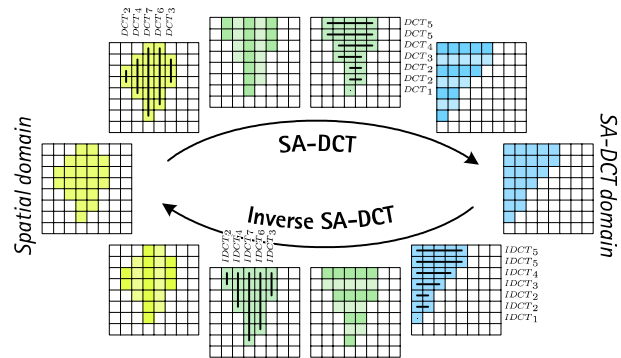


Fig. 1.　Illustration of the shape-adaptive DCT transform and its inverse. Transformation is computed by cascaded application of one-dimensional varying-length DCT transforms, along the columns and along the rows.

In this paper we present this novel approach for the denoising of grayscale as well as of color images. Extension to color images is based on a luminance-chrominance color-transformation and exploits the structural information obtained from the luminance channel to drive the shape-adaptation for the chrominance channels. Such adaptation strategy enables accurate preservation and reconstruction of image details and structures and yields estimates with a very good visual quality. Additionally, we discuss and analyze its application to deblocking and deringing of block-DCT compressed images. Particular emphasis is given to the deblocking of highly-compressed color images.

Since the SA-DCT is implemented as standard in modern MPEG hardware, the proposed techniques can be integrated within existing video platforms as a pre- or post-processing filter.

The paper is organized as follows. We begin with the considered observation model and notation. In Section III we recall the main points of the Anisotropic LPA-ICI technique. Various aspects and details of the shape-adaptive DCT transform are given in Section IV. The proposed *Pointwise SA-DCT denoising* algorithm is then introduced in Section V, which constitutes the core of the paper. The application to deblocking and deringing is given in Section VI, where we relate the quantization table with the value of the variance to be used for the filtering. In Section VII we present the extension of the proposed methods to color image filtering, describing the employed color-space transformations and the structural constraints which are imposed on the chrominances. The last section is devoted to results and discussions: we provide a comprehensive collection of experiments and comparisons which demonstrate the advanced performance of the proposed algorithms.

## II. OBSERVATION MODEL AND NOTATION

We consider noisy observations $z$ of the form

$$z(x) = y(x) + \eta(x), \quad x \in X, \qquad (1)$$

where $y$ is the original image, $\eta(x) \sim \mathcal{N}(0, \sigma^2)$ is independent Gaussian white noise, $x$ is a spatial variable belonging to the image domain $X \subset \mathbb{Z}^2$. At the beginning we restrict
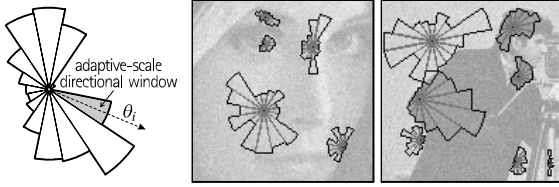
Fig. 2. Anisotropic LPA-ICI. From left to right: sectorial structure of the anisotropic neighborhood achieved by combining a number of adaptive-scale directional windows; some of these windows selected by the ICI for the noisy *Lena* and *Cameraman* images.

ourself to grayscale images (and thus scalar functions), while later (from Section VII) we consider also color images.

Given a function $f : X \to \mathbb{R}$, a subset $U \subset X$, and a function $g : U \to \mathbb{R}$, we denote by $f_{|U} : U \to \mathbb{R}$ the restriction of $f$ on $U$, $f_{|U}(x) = f(x) \ \forall x \in U$, and by $g^{|X} : X \to \mathbb{R}$ the zero-extension of $g$ to $X$, $\left(g^{|X}\right)_{|U} = g$ and $g^{|X}(x) = 0$ $\forall x \in X \setminus U$. The characteristic (indicator) function of $U$ is defined as $\chi_U = 1_{|U}^{|X}$. We denote by $|U|$ the cardinality (i.e. the number of its elements) of $U$. The symbol "⊛" stands for the convolution operation.

## III. Anisotropic LPA-ICI

The approach is based on the Intersection of Confidence Intervals (ICI) rule, a method originally developed for pointwise adaptive estimation of 1D signals [21], [24]. The idea has been generalized for 2D image processing, where adaptive-size quadrant windows have been used [25]. Significant improvement of this approach has been achieved on the basis of anisotropic directional estimation [26], [17]. Multidirectional sectorial-neighborhood estimates are calculated for every point and the ICI rule is exploited for the adaptive selection of the size of each sector. Thus, the estimator is anisotropic and the shape of its support adapts to the structures present in the image. In Figure 2 we show some examples of these anisotropic neighborhoods for the *Lena* and *Cameraman* images. The developed anisotropic estimates are highly sensitive with respect to change-points, and allow to reveal fine elements of images from noisy observations.

Let us present the Anisotropic LPA-ICI method in more detail. For every specified direction $\theta_k$, $k = 1, \dots, K$, a varying-scale family of directional-LPA convolution kernels $\{g_{h,\theta_k}\}_{h \in H}$ is used to obtain a corresponding set of directional varying-scale estimates $\{\hat{y}_{h,\theta_k}\}_{h \in H}$, $\hat{y}_{h,\theta_k} = z \circledast g_{h,\theta_k}$, $h \in H$, where $H \subset \mathbb{R}^+$ is the set of scales. These estimates are then compared according to the ICI rule, and as a result an adaptive scale $h^+(x, \theta_k) \in H$ is defined for every $x \in X$ and for every direction $\theta_k$. The corresponding adaptive-scale estimates $\hat{y}_{h^+(x,\theta_k),\theta_k}(x)$ are then "fused" together in an adaptive convex combination in order to yield the final anisotropic LPA-ICI estimate.

However, in this paper we are not interested in this anisotropic estimate. Instead, we consider only the adaptive neighborhood $U_x^+$, constructed as the union of the supports of the directional adaptive-scale kernels $g_{h^+(x,\theta_k),\theta_k}$, $U_x^+ = \bigcup_{k=1}^K \operatorname{supp} g_{h^+(x,\theta_k),\theta_k}$, which we use as the support for a shape-adaptive transform. Observe that, being convolution

kernels, the LPA kernels $g_{h,\theta_k}$ are always "centered" at the origin, therefore $U_x^+$ is a neighborhood of the origin. The actual adaptive neighborhood of $x$, which contains the observations that are used for estimation, is instead $\tilde{U}_x^+ = \{v \in X : (x - v) \in U_x^+\}$. The neighborhoods shown in Figure 2 are in fact examples of $\tilde{U}_x^+$ for a few points $x \in X$.

Let us remark that there is a substantial difference between image segmentation, in which the image is decomposed in a limited number ($\ll |X|$) of non-overlapping subsets (image segments), and the Anisotropic LPA-ICI, which for every $x \in X$ provides an adaptive neighborhood $\tilde{U}_x^+$ of $x$. In particular, because of the nonparametric nature of the procedure, neighborhoods corresponding to adjacent points do usually overlap.

## IV. Shape-Adaptive DCT transform

The SA-DCT [47], [49] is computed by cascaded application of one dimensional varying-length DCT transforms first on the columns and then on the rows that constitute the considered region. Several improvements over its original definition have been proposed. We exploit the most significant [27], which concern the normalization of the transform and the subtraction of the mean and which have a fundamental impact on the use of the SA-DCT for image filtering. Additionally, an alternative scheme for the coefficients' alignment is also utilized.

### A. Orthonormal Shape-Adaptive DCT

The normalization of the SA-DCT is obtained by normalization of the individual one-dimensional transforms that are used for transforming the columns and rows. In terms of their basis elements, they are defined as:

$$\psi_{L,m}^{\text{1D-DCT}}(n) = c_m \cos\left(\frac{\pi m (2n+1)}{2L}\right), \quad m, n = 0, \dots, L-1, \quad (2)$$
$$c_0 = \sqrt{1/L}, \ c_m = \sqrt{2/L}, \ m > 0. \quad (3)$$

Here $L$ stands for the length of the column or row to be transformed. The normalization in (2) is indeed the most natural choice, since in this way all the transforms used are orthonormal and the corresponding matrices belong to the orthogonal group. Therefore, the SA-DCT – which can be obtained by composing two orthogonal matrices – is itself an orthonormal transform. A different normalization of the 1D transforms would produce, on an arbitrary shape, a 2D transform that is non-orthogonal (for example as in [47], [49], where $c_0 = \sqrt{2}/L$ and $c_m = 2/L$ for $m > 0$).

Let us denote by $T_U : \mathcal{U} \to \mathcal{V}_U$ the orthonormal SA-DCT transform obtained for a region $U \subset X$, where $\mathcal{U} = \{f : U \to \mathbb{R}\}$ and $\mathcal{V}_U = \{\varphi : V_U \to \mathbb{R}\}$ are function spaces and $V_U \subset \mathbb{Z}^2$ indicates the domain of the transform coefficients. Let $T_U^{-1} : \mathcal{V}_U \to \mathcal{U}$ be the inverse transform of $T_U$. We indicate the thresholding (or quantization) operator as $\Upsilon$. Thus, the SA-DCT-domain processing of the observations $z$ on a region $U$ can be written as $\hat{y}_U = T_U^{-1}\left(\Upsilon\left(T_U\left(z_{|U}\right)\right)\right)$, $\hat{y}_U : U \to \mathbb{R}$. From the orthonormality of $T$ and the model (1) follows that $T_U\left(z_{|U}\right) = T_U\left(y_{|U}\right) + \bar{\eta}$, where $\bar{\eta} = T_U(\eta_{|U})$ is again Gaussian white noise with variance $\sigma^2$ and zero mean.
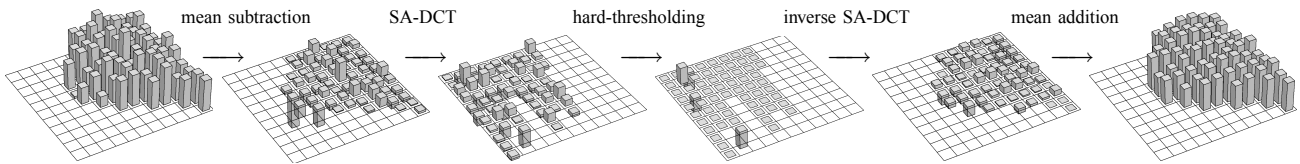
Fig. 3. Hard-thresholding in SA-DCT domain. The image data on an arbitrarily shaped region is subtracted of its mean. The zero-mean data is then transformed and thresholded. After inverse transformation, the mean is added back.

## B. Mean subtraction

There is, however, an adverse consequence of the normalization (2). Even if the signal restricted to the shape $z_{|U}$ is constant, the reconstructed $\hat{y}_U$ is usually non-constant. In [27] this behavior is termed as "mean weighting defect", and it is proposed there to attenuate its impact by applying the orthonormal SA-DCT on the zero-mean data which is obtained by subtracting from the initial data $z$ its mean. After the inversion, the mean is added back to the reconstructed signal $\hat{y}_U : U \to \mathbb{R}$:

$$\hat{y}_U = T_U^{-1}\left(\Upsilon\left(T_U\left(z_{|U} - m_U(z)\right)\right)\right) + m_U(z), \quad (4)$$

where $m_U(z) = \frac{1}{|U|}\sum_{x \in U} z(x)$ is the mean of $z$ on $U$.

Although this operation – which is termed "DC separation" – is not fully justified from the approximation theory standpoint (because $m_U(z)$ is calculated from the noisy data, and by subtracting it the noise in the coefficients is no longer white), it produces visually superior results without affecting to the objective restoration performance. The DC separation (together with a special compensation called "$\Delta$DC correction") are also considered in MPEG-4 [36].

## C. Coefficient alignment

To further improve the efficiency of the SA-DCT, it has been proposed to align the coefficients obtained after the first 1D transformation along the rows in such a way as to maximize their vertical correlation before applying the second transform along the columns. Different strategies, based on different models of the underlying signal $y$, have been suggested (e.g. [3], [2]). Although they can provide a significant improvement when the data agrees with the assumed signal's model, in practice when dealing with real data only marginal improvement can be achieved over the basic alignment used in [47], [49], where coefficients with the same index $m$ (i.e. all DC terms, all first AC terms, etc.) are aligned in the same columns, regardless of the length $L$ of the current row.

In our implementation we use the following alignment formula, denoting by $m$ and $m'$ the old (i.e. the one coming from (2)) and new coefficient index, respectively: $m' = \lfloor mL_{\max}/L \rfloor$, where $L$ is the length of the current row, $m = 0, \ldots, L-1$, $L_{\max}$ is the length of the longest row in $U$, and the $\lfloor \cdot \rfloor$ brackets indicate the rounding to the nearest integer smaller or equal to $(\cdot)$.

An illustration of the SA-DCT-domain hard-thresholding, performed according to (4) and to the above coefficient alignment formula is given in Figure 3.
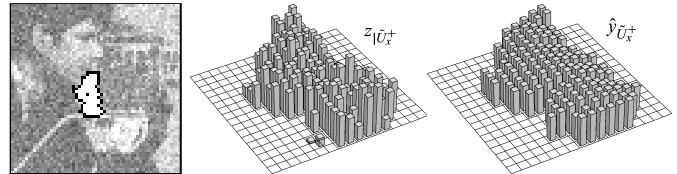


Fig. 4. From left to right: a detail of the noisy *Cameraman* showing an adaptive-shape neighborhood $\tilde{U}_x^+$ determined by the Anisotropic LPA-ICI procedure, and the image intensity corresponding to this region before and after hard-thresholding in SA-DCT domain.

## V. POINTWISE SA-DCT DENOISING

We use the anisotropic adaptive neighborhoods $\tilde{U}_x^+$ defined by the LPA-ICI as supports for the SA-DCT, as shown in Figure 4.

By demanding the local fit of a polynomial model, we are able to avoid the presence of singularities or discontinuities within the transform support. In this way, we ensure that data are represented sparsely in the transform domain, significantly improving the effectiveness of thresholding.

Before we proceed further, it is worth mentioning that the proposed approach can be interpreted as a special kind of local model selection which is adaptive with respect both to the scale and to the order of the utilized model. Adaptivity with respect to the scale is determined by the LPA-ICI, whereas the order-adaptivity is achieved by hard-thresholding. Shape-adapted orthogonal polynomials are the most obvious choice for the local transform, as they are more consistent with the polynomial modeling used to determine its support. However, in practice cosine bases are known to be more adequate for the modeling of natural images. In particular, when image processing applications are of concern, the use of computationally efficient transforms is paramount, and thus in the present paper we restrict ourself to the low-complexity SA-DCT. We refer the interested reader to [16], where our approach is considered within the more general theoretical framework of nonparametric regression.

## A. Fast implementation of the anisotropic neighborhood

In practice, we do not need a variety of different shapes as broad as in the examples of Figures 2 and 4. A much simplified neighborhood structure is used in our implementation. Narrow one-dimensional "linewise" directional LPA kernels $\{g_{h,\theta_k}\}_{h\in\{1,2,3,5,7,9\}}$ are used for $K = 8$ directions, and after the ICI-based selection of the adaptive-scales $\{h^+(x,\theta_k)\}_{k=1}^8$ the neighborhood $U_x^+$ is the octagon constructed as the polygonal hull of $\left\{\operatorname{supp} g_{h^+(x,\theta_k),\theta_k}\right\}_{k=1}^8$. Such neighborhoods are shown in Figure 5. Although the supports obtained in this
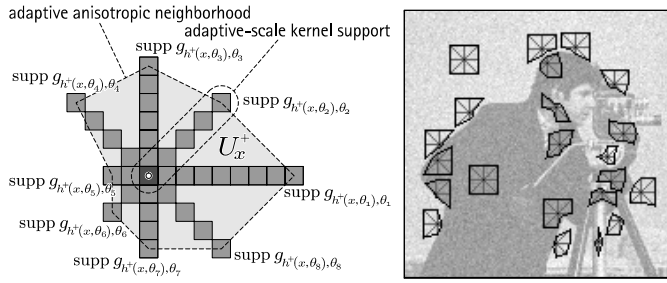
Fig. 5. Fast implementation of the LPA-ICI anisotropic neighborhoods. "Linewise" one-dimensional directional LPA kernels are used for 8 directions. The anisotropic neighborhood $U_x^+$ is constructed as the polygonal hull of the adaptive-scale kernels' supports (left). Thus, only the adaptive scales $h^+$ are needed to construct the neighborhood. Some examples of the anisotropic neighborhoods $\tilde{U}_x^+$ used for SA-DCT filtering of the noisy *Cameraman* image (right), $\sigma$=25. In our implementation we use $h \in H = \{1, 2, 3, 5, 7, 9\}$.

way have relatively simple shapes when compared to the more general examples of Figure 2, we found that this is not a significant restriction. On the contrary, a more regular boundary of the transform's support is known to improve the efficiency of the SA-DCT [4].

We note that in this particular implementation the value of the adaptive-scale $h^+(x, \theta_k)$ coincides with the length (measured in pixels) of the directional window in the direction $\theta_k$ (i.e. with the length of the support of the corresponding directional kernel).

For the sake of notation clarity, we remind that the adaptive neighborhood of $x$ used as support for the SA-DCT is $\tilde{U}_x^+$ (with tilde), which is obtained from the adaptive neighborhood $U_x^+$ (without tilde) by translation and mirroring, as defined in Section III. In both symbols the subscript "$x$" denotes the point for which the adaptive scales are obtained while the "$+$" is used to distinguish the adaptive neighborhoods from the non-adaptive ones.

### B. Local estimates

For every point $x \in X$, we construct a *local* estimate $\hat{y}_{\tilde{U}_x^+} : \tilde{U}_x^+ \to \mathbb{R}$ of the signal $y$ by thresholding in SA-DCT domain as in (4),

$$\hat{y}_{\tilde{U}_x^+} = T_{\tilde{U}_x^+}^{-1} \left( \Upsilon_x \left( \varphi_{z,x} \right) \right) + m_{\tilde{U}_x^+}(z), \qquad (5)$$

where the transform-domain coefficients $\varphi_{z,x} : V_{\tilde{U}_x^+} \to \mathbb{R}$ are calculated as

$$\varphi_{z,x} = T_{\tilde{U}_x^+} \left( z_{|\tilde{U}_x^+} - m_{\tilde{U}_x^+}(z) \right), \qquad (6)$$

and $\Upsilon_x$ is a hard-thresholding operator based on the threshold

$$\sigma \sqrt{2 \ln |\tilde{U}_x^+| + 1}. \qquad (7)$$

This threshold is essentially Donoho's "universal" threshold [9].

An estimate of the total sample variance $\text{tsvar}\{\hat{y}_{\tilde{U}_x^+}\}$ of $\hat{y}_{\tilde{U}_x^+}$ is given as sum of variances of the transform coefficients which are used for reconstruction. It has the form

$$\text{tsvar}\{\hat{y}_{\tilde{U}_x^+}\} = \sigma^2 \left( 1 + N_x^{\text{har}} \right), \qquad (8)$$

where $N_x^{\text{har}}$ is the number of non-zero coefficients after thresholding (so-called "number of harmonics") and the unit addend accounts for the addition of the mean after the inversion of the transform.

Since the anisotropic neighborhoods corresponding to nearby points are usually overlapping, and since the SA-DCT is a complete system (basis) for an individual support $\tilde{U}_x^+$, the overall approach is obviously overcomplete.

### C. Global estimate as aggregation of local estimates

In order to obtain a single *global* estimate $\hat{y} : X \to \mathbb{R}$ defined on the whole image domain, all the local estimates (5) are averaged together using adaptive weights $w_x \in \mathbb{R}$ in the following convex combination:

$$\hat{y} = \frac{\sum_{x \in X} w_x \hat{y}_{\tilde{U}_x^+}^{|X}}{\sum_{x \in X} w_x \chi_{\tilde{U}_x^+}}. \qquad (9)$$

It is a standard approach to use weights $w_x$ that are inversely proportional to the average sample variance of $\hat{y}_{\tilde{U}_x^+}$, $\text{tsvar}\{\hat{y}_{\tilde{U}_x^+}\}/|\tilde{U}_x^+|$. As shown in [22] for the case of sliding $8 \times 8$ block DCT denoising, such a simple weighting enables to attain the same performance achievable with much more involved models of the blocks' statistics.

However, this simple approach is inadequate when instead of fixed-size blocks one is considering adaptive regions with arbitrary shape and size. In particular, not only the size of the regions may vary, but also the number of overlapping shapes may be different for different points. If the inverse of the average variances are used as weights, it can be observed that when regions of significantly different sizes overlap (this may happen along edges or transitions), then the local estimates corresponding to larger regions will inevitably "submerge" the finer details restored by smaller regions.

Crucial compensation of these oversmoothing effects can be obtained by dividing the weights by the square of the size of the support, and we define $w_x$ as

$$w_x = \frac{\text{tsvar}\{\hat{y}_{\tilde{U}_x^+}\}^{-1}}{|\tilde{U}_x^+|} = \frac{\sigma^{-2}}{(1 + N_x^{\text{har}}) |\tilde{U}_x^+|}. \qquad (10)$$

Let us observe that in areas where the size of the adaptive neighborhood is nearly constant (e.g. within smooth parts of the image) the weights (10) are inversely proportional to the average and to the total sample variances of the corresponding local estimates, $w_x \propto \text{tsvar}\{\hat{y}_{\tilde{U}_x^+}\}^{-1}$. Thus, we can use the weights (10) for such areas also.

The weights $w_x$ have this form because the total sample variance $\text{tsvar}\{\hat{y}_{\tilde{U}_x^+}\}$ is obviously an upper bound for the pointwise residual-noise variance of the local estimate $\hat{y}_{\tilde{U}_x^+}$ (such pointwise variance is not necessarily uniform over $\tilde{U}_x^+$), while the extra factor $|\tilde{U}_x^+|$ addresses the correlation that exists between overlapping neighborhoods (the number of overlapping neighborhoods is loosely proportional to their size). Qualitatively speaking, these weights favour estimates which correspond to sparser representations (fewer coefficients survived thresholding, and thus lower variance) and at the same time avoid that estimates with a small support (thus representing image details) are oversmoothed by other overlapping estimates which have a large support (which usually are strongly correlated among themselves and outnumber estimates of a smaller support).

Fig. 6.    A fragment of *Lena*. From left to right: original, noisy observation ($\sigma$=25, PSNR=20.18dB), BLS-GSM estimate [42] (PSNR=31.69dB), and the proposed Pointwise SA-DCT estimate (PSNR=31.66).

### D. Wiener filtering in SA-DCT domain

Using the same approach as for thresholding, we introduce an empirical Wiener filter in the SA-DCT domain. It assumes that an estimate $\hat{y}$ of $y$ is known (in practice, we obtain this estimate using the above thresholding technique). For every $x \in X$, let $\varphi_{\hat{y},x} : V_{\tilde{U}_x^+} \to \mathbb{R}$ be the SA-DCT (on $\tilde{U}_x^+$) coefficients of $\hat{y}$ where the mean $m_{\tilde{U}_x^+}(z)$ of $z$ is subtracted before applying the transform:

$$\varphi_{\hat{y},x} = T_{\tilde{U}_x^+}\left(\hat{y}_{|\tilde{U}_x^+} - m_{\tilde{U}_x^+}(z)\right). \tag{11}$$

The *local* Wiener estimate $\hat{y}_{\tilde{U}_x^+}^{\text{wi}}$ is defined as

$$\hat{y}_{\tilde{U}_x^+}^{\text{wi}} = T_{\tilde{U}_x^+}^{-1}\left(\omega_x \varphi_{z,x}\right) + \varpi_x m_{\tilde{U}_x^+}(z), \tag{12}$$

where the SA-DCT coefficients $\varphi_{z,x}$ of $z$ are calculated as in (6), and $\omega_x \in \mathcal{V}_{\tilde{U}_x^+}$ and $\varpi_x \in \mathbb{R}$ are respectively the Wiener attenuation factors for $\varphi_{z,x}$ and for the subtracted mean value $m_{\tilde{U}_x^+}(z)$,

$$\omega_x = \frac{\varphi_{\hat{y},x}^2}{\varphi_{\hat{y},x}^2 + \sigma^2}, \qquad \varpi_x = \frac{m_{\tilde{U}_x^+}^2(\hat{y})}{m_{\tilde{U}_x^+}^2(\hat{y}) + \sigma^2/|\tilde{U}_x^+|}. \tag{13}$$

The *global* estimate $\hat{y}^{\text{wi}}$ can be obtained analogously as in (9), using the convex combination with the adaptive weights $w_x^{\text{wi}}$:

$$\hat{y}^{\text{wi}} = \frac{\sum_{x \in X} w_x^{\text{wi}} \hat{y}_{\tilde{U}_x^+}^{\text{wi}\,|X}}{\sum_{x \in X} w_x^{\text{wi}} \chi_{\tilde{U}_x^+}}, \quad w_x^{\text{wi}} = \frac{\sigma^{-2}}{(\varpi_x^2 + \sum_{V_{\tilde{U}_x^+}} \omega_x^2)|\tilde{U}_x^+|}. \tag{14}$$

Similarly to (10), the term $\sigma^2(\varpi_x^2 + \sum_{V_{\tilde{U}_x^+}} \omega_x^2)$ in the adaptive weights corresponds to an estimate of the total sample variance of $\hat{y}_{\tilde{U}_x^+}^{\text{wi}}$.

The Pointwise SA-DCT results which we present in this paper correspond to the $\hat{y}^{\text{wi}}$ estimate (14), obtained using the thresholding estimate $\hat{y}$ (9) as a reference for the calculation of the Wiener attenuation factors $\omega_x, \varpi_x$ (13).

### VI. POINTWISE SA-DCT FOR DEBLOCKING AND DERINGING OF BLOCK-DCT COMPRESSED IMAGES

The scope of the proposed filtering method is not limited to denoising only, and in this section we extend the above Pointwise SA-DCT denoising algorithm into an high-quality image deringing and deblocking filter for B-DCT compressed images.
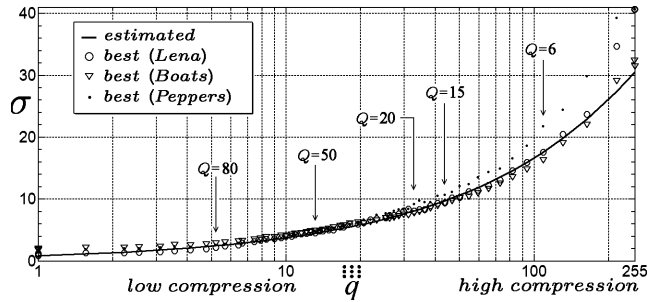


Fig. 8.    Agreement between the values of $\sigma$ estimated by Equation (15) and the best ones (found experimentally), which give the highest PSNR for the filtered *Lena*, *Boats*, and *Peppers* images.

### A. Motivation

The new wavelet-based JPEG-2000 image compression standard solved many of the drawbacks of its predecessor JPEG, which relies on the $8 \times 8$ B-DCT. The use of a wavelet transform computed globally on the whole image, as opposed to the localized B-DCT, does not introduce any blocking artifacts and allows it to achieve a very good image quality even at high compression rates. Unfortunately, this new standard has received so far only very limited endorsement from digital camera manufacturers and software developers. As a matter of fact, the classic JPEG still dominates the consumer market and the near-totality of pictures circulated on the internet is compressed using this old standard. Moreover, the B-DCT is the workhorse on which even the latest MPEG video coding standards rely upon. There are no convincing indicators suggesting that the current trend is about to change any time soon. All these facts, together with the ever growing consumer demand for high quality imaging, makes the development of advanced and efficient post-processing (deblocking, deringing, etc.) techniques still a very actual and relevant application area.

### B. Modeling

While more sophisticated models of B-DCT-domain quantization noise have been proposed by many authors, we model this degradation as some additive noise. Thus, we use the observation model $z = y + \eta$ of Equation (1), where $y$ is the original (non-compressed) image, $z$ its observation after

Fig. 7.    A fragment of *Cameraman*. From left to right: original, noisy observation ($\sigma$=25, PSNR=20.14dB), BLS-GSM estimate [42] (PSNR=28.35dB), and the proposed Pointwise SA-DCT estimate (PSNR=29.11dB).

quantization in B-DCT domain, and $\eta$ is noise with variance $\sigma^2$. In order to apply the Pointwise SA-DCT filter we need a suitable value for the variance $\sigma^2$. We estimate it directly from the quantization table $\mathbf{Q} = [q_{i,j}]_{i,j=1,\dots,8}$ using the following empirical formula:

$$\sigma^2 = 0.69 \cdot \left(\overset{\text{\tiny ::::}}{q}\right)^{1.3}, \qquad \overset{\text{\tiny ::::}}{q} = \tfrac{1}{9}\sum\nolimits_{i,j=1}^{3} q_{i,j}. \qquad (15)$$

This formula uses only the mean value $\overset{\text{\tiny ::::}}{q}$ of the nine table entries which correspond to the lowest-frequency DCT harmonics (including the DC-term) and has been experimentally verified to be adequate for a wide range of different quantization tables and images. In Figure 8 we show how the values of $\sigma$ calculated by Equation (15) agree with the best values found experimentally for the *Lena*, *Boats*, and *Peppers* images compressed with different quantization tables corresponding to JPEG with quality $Q$=1,...,100 (and thus $\overset{\text{\tiny ::::}}{q}$=1,...,255). Note that a higher compression (e.g. JPEG with small $Q$) corresponds to a larger value for this variance (i.e. $Q$ and $\overset{\text{\tiny ::::}}{q}$ are inversely related). The standard-deviation $\sigma$ is not linear with respect to the $q_{i,j}$'s, a fact which reflects the non-uniformity of the distribution of the B-DCT coefficients.

Note that the $\sigma^2$ which is calculated by (15) is *not* an estimate of the variance of compressed image, *nor* it is an estimate of the variance of the difference between original and compressed images. Instead, it is simply the assumed value for the variance of $\eta$ in the observation model (1). Roughly speaking, it is the variance of some hypothetical noise which, if added to the original image $y$, would require – in order to be removed – the same level of adaptive smoothing which is necessary to suppress the artifacts generated by the B-DCT quantization with the table $\mathbf{Q}$. Much larger or much smaller values of $\sigma^2$ would respectively result in oversmoothing or leave the compression artifacts unfiltered.

Figures 9 and 10 show fragments of the JPEG-compressed grayscale *Cameraman* image obtained for two different compression levels (JPEG quality $Q$=6 and $Q$=15) and the corresponding Pointwise SA-DCT filtered estimates. For these two cases the estimated standard-deviations are $\sigma$=17.6 and $\sigma$=9.7.

Let us observe that the procedure defined by (15) can be used in a straightforward manner, because the quantization tables are always (and necessarily) either provided with the coded data, or fixed in advance by the compression standard. It allows to apply the Pointwise SA-DCT denoising algorithm



Fig. 9.    Details of the JPEG-compressed *Cameraman* ($Q$=6, bpp=0.19, PSNR=25.03dB) and of the corresponding Pointwise SA-DCT estimate (PSNR=26.11dB). The estimated standard deviation for this highly compressed image is $\sigma$=17.6.



Fig. 10.    Details of the JPEG-compressed *Cameraman* ($Q$=15, bpp=0.37, PSNR=27.71dB) and of the corresponding Pointwise SA-DCT estimate (PSNR=28.58dB). The estimated standard deviation for this compressed image is $\sigma$=9.7.

of Section V as an effective deblocking and deringing filter for B-DCT coded images and videos. The proposed method is particularly relevant for video postprocessing, since it can exploit the SA-DCT hardware of MPEG-4 decoders.

## VII. POINTWISE SA-DCT FILTERING OF COLOR IMAGES WITH STRUCTURAL CONSTRAINT IN LUMINANCE-CHROMINANCE SPACE

The extension from grayscale to color images of our denoising and deblocking approach is based on a very simple, yet powerful strategy. The key idea is the following: the structures (e.g. objects, edges, details, etc.) which determine the adaptive shapes are the same across all three color channels, thus the same shapes should be used for the SA-DCT filtering of

the three channels. In order to increase its effectiveness, the method is implemented after transformation in a luminance-chrominance color-space. We call it *structural constraint in luminance-chrominance space* and it fully exploits the shape-adaptive nature of our approach without adding anything to its complexity.

### A. Luminance-chrominance space

We generalize the observation model (1) to color data. Let $y = [y_R \; y_G \; y_B]$ be the original color image in the *RGB* color space. We consider noisy observations $z = [z_R \; z_G \; z_B]$ of the form

$$z_C = y_C + \eta_C, \qquad C = R, G, B, \qquad (16)$$

where the noise $\eta = [\eta_R \; \eta_G \; \eta_B]$ is independent Gaussian, $\eta_C(\cdot) \sim \mathcal{N}(0, \sigma_C^2)$, $C = R, G, B$.

In order to deal with color images, we first perform a color-space transformation, aiming at reducing the strong correlation between channels which is typical of the *RGB* space. In particular, we consider the "*opponent*" and the *YUV/YCbCr* color spaces [41]. Up to some normalization, the transformation to these color spaces can be expressed by multiplication of a column vector with the $R$, $G$, and $B$ components against one of the matrices

$$\mathbf{A}_{opp} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{\sqrt{6}} & 0 & \frac{-1}{\sqrt{6}} \\ \frac{1}{3\sqrt{2}} & \frac{-\sqrt{2}}{3} & \frac{1}{3\sqrt{2}} \end{bmatrix}, \quad \mathbf{A}_{yuv} = \begin{bmatrix} 0.30 & 0.59 & 0.11 \\ -0.17 & -0.33 & 0.50 \\ 0.50 & -0.42 & -0.08 \end{bmatrix}.$$

Although purists may consider it an abuse of terminology, we call "luminance" and "chrominances" not only the components the *YUV* space, but also those of the opponent color space. We denote the luminance channel as $Y$, and the chrominances as $U$ and $V$.

In such luminance-chrominance decompositions, the original inter-channel correlation of the *RGB* space is captured into the luminance channel, which thus enjoys a better signal-to-noise ratio (SNR), whereas the chrominance channels contain the differential information among the *RGB* channels.

We then come to the following observation model in luminance-chrominance space,

$$z_C = y_C + \eta_C, \qquad C = Y, U, V, \qquad (17)$$

where $[z_Y \; z_U \; z_V] = [z_R \; z_G \; z_B] \mathbf{A}^T$, $[y_Y \; y_U \; y_V] = [y_R \; y_G \; y_B] \mathbf{A}^T$, and $\eta_C(\cdot) \sim \mathcal{N}(0, \sigma_C^2)$, $C = Y, U, V$.

Ideally, the $Y$, $U$, and $V$ channels are considered as independent. Therefore, the common approach for color denoising in luminance-chrominance space is to filter the three channels (i.e. $z_Y$, $z_U$, and $z_V$) separately and independently one from the other.

However, when considering natural images, the different color channels always share some common features which are inherited from the structures and from the objects depicted in the original image. In particular, it can be observed that along the objects' boundaries all color channels usually exhibit some simultaneous discontinuities or sharp transitions.

We exploit this kind of structural correlation by imposing that the three transform's supports which are used for the filtering of $z_Y$, $z_U$, and $z_V$ at a particular location have the same adaptive shape. In practice, we use for all three channels
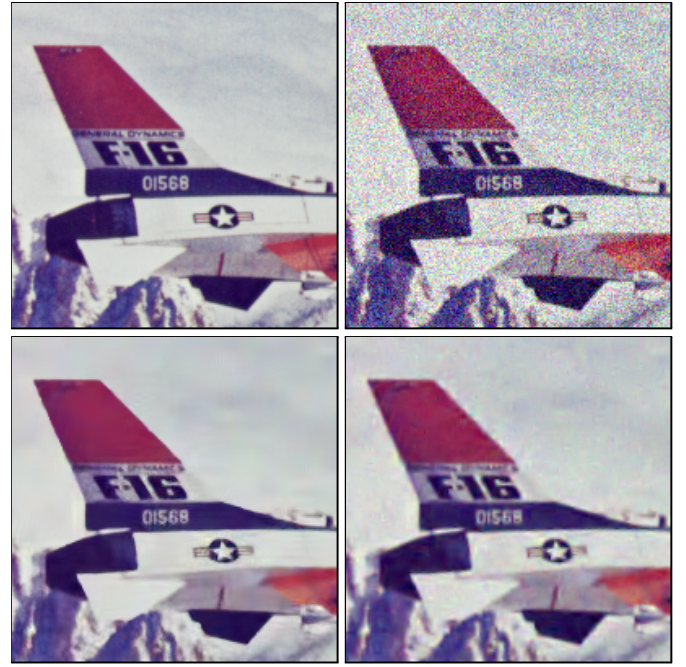


Fig. 11.    Fragments of the original *F-16* image (top-left), of its noisy observation ($\sigma$=30, PSNR=18.59dB) (top-right) and of two denoised estimates: our Pointwise SA-DCT estimate (bottom-left), and the ProbShrink-MB [40] estimate (bottom-right). The PSNR for the two estimates is 31.59dB and 30.50dB, respectively.

the adaptive neighborhoods defined by the Anisotropic LPA-ICI for the luminance channel.

Such a constraint has the effect that whenever some structure is detected (in the luminance channel by the LPA-ICI), it is taken into account (and thus preserved) for the filtering of all three channels. Restricted to the adaptive supports, however, the channels are assumed as independent, and thus the transform-domain hard-thresholding and Wiener filtering are still performed for each channel independently from the others.

After the filtering of the three channels, inverse color-transformation returns the estimate of the original image $y$ in the *RGB* space.

### B. Pointwise SA-DCT denoising in luminance-chrominance space

The noise variances for the $Y$, $U$, and $V$ channels can be calculated as the elements of the vector $[\sigma_Y^2 \; \sigma_U^2 \; \sigma_V^2] = [\sigma_R^2 \; \sigma_G^2 \; \sigma_B^2] \mathbf{A}^{T2}$, where $\sigma_R^2$, $\sigma_G^2$, and $\sigma_B^2$ are the noise variances for the $R$, $G$, and $B$ channels and $\mathbf{A}^{T2}$ is the transposed color transformation matrix with all elements squared. For denoising, the opponent color transformation is preferable because of the orthogonality of the rows of $\mathbf{A}_{opp}$.

The better SNR of the luminance and its higher "information content" are the two main reasons why it is in this channel that we look for structures. There are also other reasons. In natural images it often happens that uniformly colored objects present luminance variations due to non-uniform illumination or shadowing: such transitions cannot be detected from the chrominances. On the other hand, it is quite rare that abrupt changes appear in the chrominances and not in the luminance.

Fig. 12. Fragments of the JPEG-compressed ($Q$=10, 0.25bpp, PSNR=26.87dB), and restored *F-16* color image (PSNR=28.30dB) using the proposed Pointwise SA-DCT deblocking filter in luminance-chrominance space.

Therefore, it is sufficient to perform the LPA-ICI adaptive-scale selection on the luminance channel only.

### C. Deblocking and deringing of B-DCT compressed color images

The proposed strategy for color image filtering is also particularly effective for deblocking and deringing color images.

When compressing color images or video, the standard approach (e.g. in the JPEG and MPEG), is to first perform the *YUV* color transformation and then compress the resulting three channels separately. According to the modeling in the previous sections, we assume that the original (non-compressed) image $y$ in the *RGB* color space is represented, after B-DCT quantization in *YUV* space, as the $z_C$ in the observation model (17), where $y_Y$, $y_U$ and $y_V$ are the luminance and chrominance channels of $y$, and $z_Y$, $z_U$ and $z_V$ are the corresponding channels after quantization in B-DCT domain.

We estimate the variances $\sigma_Y^2$, $\sigma_U^2$, and $\sigma_V^2$ of $\eta_C$, $C = Y, U, V$, from the corresponding quantization tables for the luminance and chrominance channels, using formula (15). However, if (as it is commonly done) the chrominance channels are downsampled, then the estimated variances for the chrominances need to be further multiplied by 2, in order to account for the coarser sampling.

Usually, the quantization tables $\mathbf{Q}^U$ and $\mathbf{Q}^V$ used for the two chrominances coincide, $\mathbf{Q}^U = \mathbf{Q}^V = \mathbf{Q}^{UV}$. Following standard models of the human visual system, a higher compression is typically performed on the chrominances than on the luminance. Thus, it is typical that the estimated variances are such that $2\sigma_Y^2 < \sigma_U^2 = \sigma_V^2$. Even at relatively high bit-rates, the compression of the chrominance channels can be quite aggressive.

As for color image denoising, we approach color data in a channel-by-channel manner imposing a unique structural constraint among the three channels. This allows to filter the chrominance channels restoring the structural information which was lost due to quantization and coarse sampling. The peculiarity of our approach is easily explained and demonstrated through the next example.

Figures 12 and 13 present a very common scenario. It can be seen that only very few AC-terms of the chrominance blocks survive to quantization, and the resulting chrominance channels end up with the vast majority of blocks represented
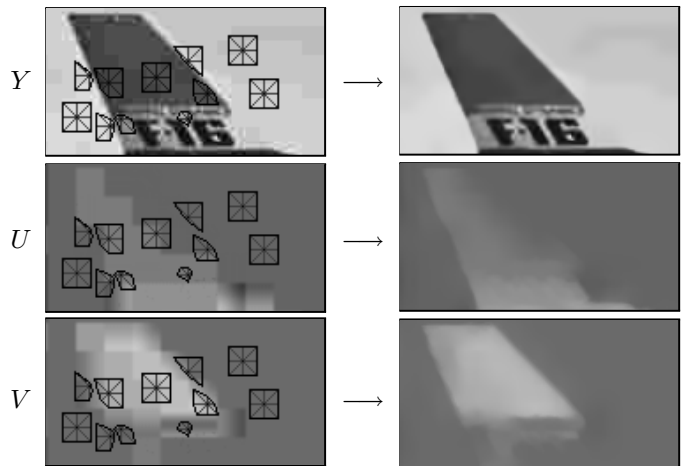


Fig. 13. The adaptive anisotropic neighborhoods are selected by the LPA-ICI on the luminance channel (left-top). Observe that the neighborhoods are not affected by the blocking artifacts and yet are quite accurate with respect to the image features. These neighborhoods are used for SA-DCT filtering of the luminance as well as of the two chrominances (left-middle and left-bottom). The result of such filtering is shown in the right column. The color estimate obtained after inverse *YUV* color-transformation is shown in Figure 12.

by the DC-term only. It results in unpleasant color-bleeding artifacts along edges between differently colored objects. At the same time, on smoother areas the uneven hue due to quantization becomes particularly noticeable. In this example, the values of $\sigma_Y$ and $\sigma_U = \sigma_V$ calculated according to formula (15) are 12.6 and 27.1, respectively.

As shown in Figure 13(left), we use for all three channels the adaptive neighborhoods defined by the Anisotropic LPA-ICI for the $Y$ channel, because it is in the luminance that the structural information is usually better preserved after compression.

Figure 13(right) shows that the proposed method effectively attenuates ringing and blocking artifacts, faithfully preserving the structures and the salient feature in the image. Moreover, it demonstrates its ability of *reconstructing the missing structural information* in the chrominance channels, where the details of the tail of the plane are clearly revealed, with precise boundaries. The obtained color estimate, shown in Figure 12(right), is then quite sharp, with well-defined edges, and the color-bleeding artifacts (clearly visible in the JPEG-compressed image) are accurately corrected.

## VIII. EXPERIMENTS AND RESULTS

We conclude the paper with a number of experimental results and comparisons which demonstrate the state-of-the-art performance of the developed algorithms.

### A. Grayscale denoising

Let us start with filtering of grayscale images corrupted by additive Gaussian white noise. In Table I we compare our results against those reported by other authors. In terms of PSNR, the results of our estimates are high, often outperforming all other methods. Additional results are given in Table II for more images and levels of noise. We emphasize the outstanding preservation of sharp details which we demonstrate in Figures 6, 7, and 14, while almost no visible artifacts

TABLE I

PSNR (dB) COMPARISON TABLE FOR THE DENOISING OF THE GRAYSCALE *Lena*, *Boats*, *Peppers*, AND *House* TEST IMAGES WITH DIFFERENT LEVELS OF GAUSSIAN NOISE.

| Method      $\sigma$ | *Lena* 512×512 | | | *Boats* 512×512 | | | *Peppers* 256×256 | | | *House* 256×256 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 15 | 20 | 25 | 15 | 20 | 25 | 15 | 20 | 25 | 15 | 20 | 25 |
| Pointwise SA-DCT | 33.86 | 32.62 | 31.66 | 31.79 | 30.49 | 29.47 | 32.44 | 31.04 | 29.92 | 34.14 | 32.92 | 31.92 |
| BLS-GSM (Portilla et al.) [42] | 33.90 | 32.66 | 31.69 | 31.70 | 30.38 | 29.37 | 31.74 | 30.31 | 29.21 | 33.63 | 32.39 | 31.40 |
| Patch-based (Kervrann and Boulanger) [29] | 33.70 | 32.64 | 31.73 | 31.44 | 30.12 | 29.20 | 32.13 | 30.59 | 29.73 | 34.08 | 32.90 | 32.22 |
| MGGD (Cho and Bui) [7] | 33.70 | 32.46 | 31.48 | 31.46 | 30.14 | 29.12 | – | – | – | – | – | – |
| Recursive Anisotropic LPA-ICI [17], [18] | 32.72 | 31.44 | 30.43 | 30.87 | 29.58 | 28.58 | 31.78 | 30.30 | 29.16 | 33.18 | 31.82 | 30.73 |

TABLE II

GRAYSCALE IMAGE DENOISING PERFORMANCE AS PSNR (dB) FOR THE PROPOSED POINTWISE SA-DCT ALGORITHM.

| $\sigma$ | *Lena* | *Peppers* | *Barbara* | *C.man* | *House* | *Hill* |
|---|---|---|---|---|---|---|
| 5 | 38.54 | 37.99 | 37.47 | 38.15 | 39.38 | 37.03 |
| 10 | 35.58 | 34.46 | 33.48 | 33.98 | 35.98 | 33.43 |
| 15 | 33.86 | 32.44 | 31.37 | 31.70 | 34.14 | 31.60 |
| 20 | 32.62 | 31.04 | 30.00 | 30.18 | 32.92 | 30.39 |
| 25 | 31.66 | 29.92 | 28.95 | 29.11 | 31.92 | 29.50 |
| 30 | 30.86 | 29.03 | 28.10 | 28.24 | 31.10 | 28.80 |
| 35 | 30.17 | 28.26 | 27.35 | 27.51 | 30.39 | 28.22 |
| 50 | 28.60 | 26.55 | 25.44 | 25.88 | 28.67 | 26.85 |

TABLE III

COLOR IMAGE DENOISING PERFORMANCE AS PSNR (dB) FOR THE PROPOSED POINTWISE SA-DCT ALGORITHM.

| $\sigma$ | *Lena* | *Peppers* | *Baboon* | *House* | *F-16* | *Lake* |
|---|---|---|---|---|---|---|
| 10 | 34.95 | 33.70 | 30.62 | 35.67 | 36.41 | 32.34 |
| 15 | 33.58 | 32.42 | 28.33 | 34.09 | 34.67 | 30.52 |
| 20 | 32.61 | 31.57 | 26.89 | 32.97 | 33.41 | 29.40 |
| 25 | 31.85 | 30.90 | 25.86 | 32.12 | 32.42 | 28.58 |
| 30 | 31.21 | 30.33 | 25.07 | 31.39 | 31.59 | 27.93 |
| 35 | 30.65 | 29.81 | 24.44 | 30.74 | 30.88 | 27.38 |
| 50 | 29.27 | 28.53 | 23.03 | 29.13 | 29.19 | 26.10 |
| 75 | 27.77 | 27.07 | 21.46 | 27.39 | 27.43 | 24.68 |



Fig. 14. Fragments of the Pointwise SA-DCT estimates of the *Boats*, *Peppers*, and *House* images, $\sigma$=25.

are present. Other transform-based estimates, such as those from [42], often display noticeable overshootings on the edges and unpleasant spurious oscillations. These artifacts, which are characteristic of all oscillatory transforms (including the SA-DCT), do not appear in our estimates thanks to the adaptive selection of the transform support.

### B. Color denoising

For the color denoising experiments, the variance of the additive Gaussian noise is set to be the same for all *RGB* color channels, $\sigma_R^2 = \sigma_G^2 = \sigma_B^2 = \sigma^2$. Filtering is performed after transformation to the opponent color space. Table III gives the PSNR results for the denoising of the *Lena*, *Peppers*, *Baboon*, *House*, *F-16*, and *Lake* color test-images over a wide range of values of $\sigma$.

In Table IV we compare our results against those by other state-of-the-art methods, as reported in [40]. In particular, the vector-based minimum-mean-squared-error estimator (VMMSE) [45], the multiband wavelet thresholding (MBT) [44], and the ProbShrink-multiband wavelet algorithm [40] are considered for comparison.

Let us note that the reference methods which are included in the Table IV are all multiband or vector methods, which are specifically designed for the denoising of color or multispectral

images. Such algorithms simultaneously filter all channels, exploiting the possible inter-channel correlation, and are thus inherently superior to the simpler strategy where a scalar (grayscale) denoising filter is used independently for each separate channel.

We remark that although in our approach the adaptive supports for the SA-DCT at a particular location are the same for all three channels, the SA-DCT-domain filtering is performed for each channel independently. Nevertheless, our results are very competitive and the comparison table shows that in fact the proposed technique outperforms all other reference methods.

Similarly to the grayscale case, the denoised color estimates produced by our adaptive algorithm are visually very good. A close inspection to Figures 11 and 15 may reveal the outstanding preservation of sharp details achieved by the shape-adaptive transform. At the same time, almost no visible artifacts (such as blurriness or overshootings) are present.

### C. Deblocking and deringing

In order to assess the filtering performance of the proposed method, extensive simulation experiments were performed for different types of quantization tables, several level of compression, and for grayscale as well as for color images. We reproduce the same experimental settings used by other authors and we present comparative numerical results collected in three separate tables. The first two tables contain results for grayscale images obtained using three particular quantization tables found in the literature (Table V) and using the standard JPEG (Table VI). The third and last table is dedicated to experiments with JPEG compression of color images.

Three quantization tables, usually denoted as $\mathbf{Q}1$, $\mathbf{Q}2$, and $\mathbf{Q}3$, have been used by many authors (e.g. [33] and references

Fig. 15. Fragments from the noisy ($\sigma$=25, PSNR=20.18dB) and denoised color *Peppers* image (PSNR=30.90dB), obtained using the proposed Pointwise SA-DCT algorithm.

TABLE IV
PSNR (dB) COMPARISON TABLE FOR THE DENOISING OF THE *Lena*, *Peppers*, AND *Baboon* COLOR TEST IMAGES WITH DIFFERENT LEVELS OF GAUSSIAN NOISE.

| Method | $\sigma$ | Color *Lena* 512×512 | | | | Color *Peppers* 512×512 | | | | Color *Baboon* 512×512 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 10 | 15 | 20 | 25 | 10 | 15 | 20 | 25 | 10 | 15 | 20 | 25 |
| Pointwise SA-DCT | | 34.95 | 33.58 | 32.61 | 31.85 | 33.70 | 32.42 | 31.57 | 30.90 | 30.62 | 28.33 | 26.89 | 25.86 |
| ProbShrink-MB (Pizurica et al.) [40] | | 34.60 | 33.03 | 31.92 | 31.04 | 33.44 | 32.05 | 31.12 | 30.35 | 30.17 | 27.83 | 26.38 | 25.27 |
| VMMSE (Scheunders and Driesen) [45] | | 34.02 | 31.89 | 30.24 | 28.88 | 33.12 | 31.13 | 29.67 | 28.45 | 30.68 | 28.24 | 26.63 | 25.36 |
| MBT (Scheunders) [44] | | 33.84 | 32.29 | 31.14 | 30.15 | 31.19 | 30.22 | 29.45 | 28.77 | 28.50 | 26.78 | 25.53 | 24.56 |

therein) in order to simulate various types of B-DCT compression. To help the reader identifying the considered quantization tables, we report here the first row of each table:

$\mathbf{Q}$1 $(1\cdots8,1)$ = [50 60 70 70 90 120 255 255],

$\mathbf{Q}$2 $(1\cdots8,1)$ = [86 59 54 86 129 216 255 255],

$\mathbf{Q}$3 $(1\cdots8,1)$ = [110 130 150 192 255 255 255 255].

The values of the standard deviation $\sigma$ corresponding to these three tables – calculated using formula (15) – are 12.62, 13.21, and 22.73, respectively. In terms of image degradation, they correspond to a medium to high compression level, similar to what can be obtained by using JPEG with quality $Q$=11, $Q$=9, and $Q$=5, respectively.

In Table V we present results for deblocking from B-DCT quantization performed using these specific quantization tables. We compare the results obtained by our SA-DCT algorithm against the best results obtained by any of the methods [33], [23], [35], [39], [56], [54], as reported in [33]. The results are in favor of our proposed technique, which consistently outperforms all other methods.

Further positive results are shown in Table VI for the case of deblocking from JPEG-compression. In this second table we compare against the best result obtained by any of the methods [1], [6], [43], [55], [34], [35], as reported in [1]. Also in this comparison, the SA-DCT method is found to be superior to all other techniques, outperforming them of about 0.5dB in all experiments.

In Table VII we show results for the SA-DCT filtering of JPEG-compressed color images, from very high ($Q$=4) to very low ($Q$=75) compression levels. It can be seen that the improvement is significant especially for very high and moderate compression levels. For very low compression levels (for which the compression artifacts are barely visible and thus there is typically no need for postprocessing) the improvement is still substantial for those images which present some structures or edges.

For the simulations in Table VI and Table VII as well as for all JPEG experiments presented in this paper, we use the baseline IJG JPEG implementation. For a JPEG-quality parameter $Q$=50, the top rows of the quantization tables for the luminance and chrominance channels are

$\mathbf{Q}_{Q=50}^{Y}$ $(1\cdots8,1)$ = [16 11 10 16 24 40 51 61],

$\mathbf{Q}_{Q=50}^{UV}$ $(1\cdots8,1)$ = [17 18 24 47 99 99 99 99],

and the corresponding estimated standard-deviations according to (15) are $\sigma_Y = 4.4$ and $\sigma_U = \sigma_V = 9.7$.

We conclude with two examples which highlight the very special reconstruction ability enabled by our structural constraint in luminance-chrominance space.

Figure 16(left) shows a fragment of the JPEG compressed Lena image ($Q$=20, 0.38bpp, PSNR=29.83dB). The corresponding $U$ and $V$ chrominance channels are shown in Figure 17. One can barely recognize the salient features of the image, such as the border of the hat or the contours of the eyes and nose. These structures can be faithfully restored by the use of adaptive-shape supports which are determined from the luminance channel, as shown in Figure 18. It is remarkable that even small details such as the iris can be accurately reconstructed from the coarse available information using adaptive transform's supports. The restored color image

TABLE V
PSNR (DB) COMPARISON TABLE FOR RESTORATION FROM B-DCT QUANTIZATION FOR THREE DIFFERENT QUANTIZATION MATRICES.
THE VALUES UNDER "OTHERS" CORRESPOND TO THE BEST RESULTS OF ANY OF THE METHODS [33], [23], [35], [39], [56], [54], AS
REPORTED IN [33].

| Table | *Lena* 512×512 | | | *Peppers* 512×512 | | | *Barbara* 512×512 | | |
|---|---|---|---|---|---|---|---|---|---|
| | image | others | P. SA-DCT | image | others | P. SA-DCT | image | others | P. SA-DCT |
| **Q1** | 30.70 | 31.63 | 32.12 | 30.42 | 31.33 | 32.02 | 25.94 | 26.64 | 26.79 |
| **Q2** | 30.09 | 31.19 | 31.56 | 29.82 | 30.97 | 31.45 | 25.59 | 26.32 | 26.45 |
| **Q3** | 27.38 | 28.65 | 29.03 | 27.22 | 28.55 | 29.13 | 24.03 | 24.73 | 25.13 |

TABLE VI
PSNR (DB) COMPARISON TABLE FOR RESTORATION FROM JPEG COMPRESSION OF GRAYSCALE IMAGES. THE VALUES UNDER
"OTHERS" CORRESPOND TO THE BEST RESULT OBTAINED BY ANY OF THE METHODS [1], [6], [43], [55], [34], [35], AS REPORTED IN
[1].

| Qual. | *Lena* 512×512 | | | *"Green Peppers"*[1] 512×512 | | | *Barbara* 512×512 | | |
|---|---|---|---|---|---|---|---|---|---|
| | JPEG (bpp) | | others | P. SA-DCT | JPEG (bpp) | | others | P. SA-DCT | JPEG (bpp) | | others | P. SA-DCT |
| 4 | 26.46 | (0.11) | 27.63 | 28.08 | 25.61 | (0.14) | 26.72 | 27.41 | 23.48 | (0.14) | 24.13 | 24.65 |
| 6 | 28.24 | (0.15) | 29.22 | 29.87 | 27.32 | (0.18) | 28.22 | 28.97 | 24.50 | (0.18) | 25.08 | 25.51 |
| 8 | 29.47 | (0.18) | 30.37 | 30.99 | 28.40 | (0.22) | 29.28 | 29.90 | 25.19 | (0.23) | 25.71 | 26.11 |
| 10 | 30.41 | (0.22) | 31.17 | 31.84 | 29.16 | (0.25) | 29.94 | 30.51 | 25.79 | (0.28) | 26.27 | 26.61 |
| 12 | 31.09 | (0.25) | 31.79 | 32.48 | 29.78 | (0.28) | 30.47 | 31.00 | 26.33 | (0.32) | 26.81 | 27.10 |

TABLE VII
PSNR (DB) RESULTS FOR THE POINTWISE SA-DCT FILTERING OF JPEG-COMPRESSED COLOR IMAGES. RESULTS ARE GIVEN ALSO IN
TERMS OF IMPROVEMENT-IN-*SNR* (*ISNR*, DB).

| Qual. | Color *Lena* 512×512 | | | Color *Peppers* 512×512 | | | Color *Baboon* 512×512 | | | Color *House* 256×256 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | JPEG (bpp) | P. SA-DCT | ISNR | JPEG (bpp) | P. SA-DCT | ISNR | JPEG (bpp) | P. SA-DCT | ISNR | JPEG (bpp) | P. SA-DCT | ISNR |
| 4 | 23.34(0.12) | 24.79 | 1.45 | 22.32(0.13) | 23.77 | 1.46 | 19.28(0.17) | 20.00 | 0.72 | 22.63(0.15) | 23.76 | 1.13 |
| 6 | 25.52(0.16) | 27.09 | 1.57 | 23.99(0.17) | 25.54 | 1.54 | 20.38(0.26) | 21.05 | 0.67 | 24.41(0.19) | 25.66 | 1.24 |
| 8 | 26.64(0.19) | 28.16 | 1.52 | 24.99(0.21) | 26.40 | 1.41 | 21.12(0.35) | 21.71 | 0.59 | 25.16(0.24) | 26.41 | 1.25 |
| 10 | 27.53(0.23) | 29.06 | 1.53 | 25.77(0.25) | 27.11 | 1.34 | 21.63(0.43) | 22.13 | 0.50 | 26.25(0.27) | 27.54 | 1.29 |
| 15 | 28.97(0.31) | 30.33 | 1.35 | 26.88(0.33) | 27.99 | 1.11 | 22.49(0.62) | 22.88 | 0.38 | 27.52(0.34) | 28.66 | 1.14 |
| 20 | 29.83(0.38) | 31.00 | 1.17 | 27.57(0.40) | 28.53 | 0.96 | 23.07(0.77) | 23.37 | 0.31 | 27.87(0.41) | 28.75 | 0.88 |
| 25 | 30.44(0.44) | 31.46 | 1.02 | 28.04(0.47) | 28.90 | 0.86 | 23.50(0.92) | 23.75 | 0.25 | 28.55(0.47) | 29.44 | 0.89 |
| 30 | 30.91(0.50) | 31.79 | 0.88 | 28.40(0.54) | 29.14 | 0.74 | 23.85(1.05) | 24.06 | 0.21 | 28.96(0.54) | 29.76 | 0.80 |
| 40 | 31.54(0.61) | 32.26 | 0.72 | 28.83(0.66) | 29.45 | 0.62 | 24.40(1.29) | 24.56 | 0.16 | 29.51(0.65) | 30.20 | 0.69 |
| 50 | 32.02(0.72) | 32.63 | 0.61 | 29.25(0.78) | 29.81 | 0.56 | 24.85(1.51) | 24.97 | 0.12 | 29.80(0.76) | 30.40 | 0.60 |
| 75 | 33.21(1.13) | 33.56 | 0.35 | 30.29(1.23) | 30.67 | 0.52 | 26.21(2.33) | 26.25 | 0.04 | 31.44(1.18) | 32.00 | 0.56 |

(PSNR=31.00dB) is shown in Figure 16(right). The ringing and the blocking artifacts disappeared, whereas no details have been oversmoothed, demonstrating the superior adaptivity of the approach. Moreover, thanks to the accurate reconstruction of the structures in the chrominance channels, our estimate does not exhibit any significant chromatic distortion and has a natural appearance.

Although it is well-established that the human visual system is less sensitive to distortions in the chrominances than to those in the luminance, the importance of restoring the chrominances must not be overlooked. In fact, all modern image and video compression standards are designed to exploit the characteristics of the human visual system, and thus adjust the compression rate for the luminance and chrominance channels in such a way to balance the perceptual impact of the distortions among the three channels. Therefore, when visual quality is of concern, the restoration of the different channels deserves equal attention. The downsampling and the coarser quantization of the chrominances makes their accurate restoration a much more difficult and challenging task.

Figure 19 provides a final example of the accuracy of the proposed method. First, one can see the sharp reconstruction of contours (e.g. in the legs, shoulders and head). Color-bleeding and blocking artifacts are completely suppressed, not only on smooth regions but even on rather thin details such as the snorkel. Second, the figure shows that the method is still reliable even when no useful structural information can be extracted from the luminance channel. In particular, it can be seen that the swimsuit is composed of three differently colored patches, all of which have the same luminance. This makes impossible to reconstruct the boundaries between these patches in a very sharp manner, as the only information available lies in the chrominances. Nevertheless, because the SA-DCT is a basis (complete system), the different colors of these patches are well preserved, while the transform-domain thresholding effectively suppresses the blockiness.

---

[1]In order to replicate the experiments as in [1], the "*Peppers*" image used for Table VI is the green channel of the *RGB* color *Peppers*. Let us note, however, that far more often in the literature the grayscale *Peppers* are found as the luminance channel $Y$ of the *RGB Peppers* image.
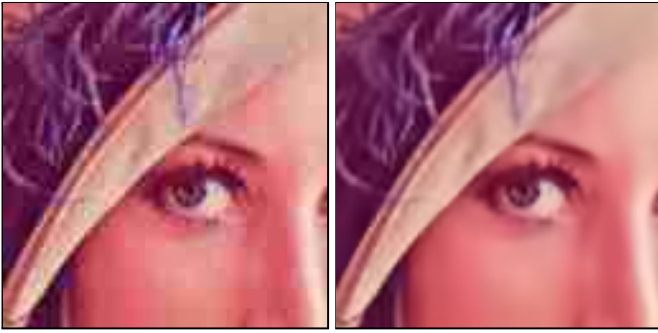
Fig. 16. Fragment of the JPEG-compressed *Lena* image ($Q$=20, 0.38bpp, PSNR=29.83dB) and of its Pointwise SA-DCT filtered estimate (PSNR=31.00dB).

## D. Subjective perceptual quality

Although the figures presented in this paper already show that the Pointwise SA-DCT estimates are visually very good, we wish to mention that an independent research as been conducted [51], [52], aiming at evaluating the perceptual quality of the estimates of many state-of-the-art denoising methods. The research was based on an extensive psychovisual experiment where several noisy images were denoised by various algorithms and then subjectively judged by a large group of observers. Aspects such as blurriness, noisiness, presence of artifacts, etc., were investigated. Our Pointwise SA-DCT algorithm was among the considered denoising algorithms, which included many state-of-the-art methods such as [42], [46], and [40]. The analysis of the results of this experimental research attest that our Pointwise SA-DCT estimates clearly
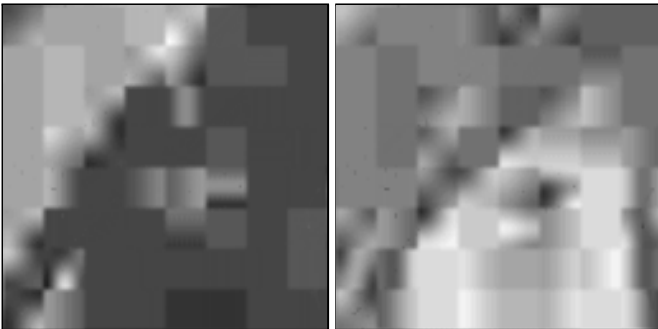


Fig. 19. Fragments of the original, compressed, and restored *Kodak image 12*. Top to bottom row: *RGB* color, luminance *Y* channel, chrominance *V* channel. From left to right: original image, JPEG-compressed ($Q$=16, 0.25bpp, PSNR=30.45), restored by proposed Pointwise SA-DCT method (PSNR=31.45).

outperformed in terms of overall subjective quality all estimates produced by techniques of other authors. In particular, it is shown that even in those cases where in terms of PSNR the Pointwise SA-DCT estimate might be inferior to the BLS-GSM estimate [42], thanks to the absence of artifacts and the better preservation of details the Pointwise SA-DCT estimate still provides a significantly superior visual quality.

## E. Complexity

When considering the computational complexity of the proposed algorithm, the first thing to observe is that the LPA-ICI technique is fast, because it is based on convolutions against one-dimensional kernels for a very limited number of directions. It constitutes a negligible computational overhead for the whole Pointwise SA-DCT filtering algorithm, whose complexity is instead essentially determined by the calculation of the forward and inverse SA-DCT transforms for every processed neighborhood.

The complexity of the algorithm is linear with respect to the size of the image and depends on the shapes of the transform supports. Since such shapes are pointwise-adaptive, the complexity depends on the particular image and noise. On average, and without resorting to fast algorithms, the asymptotic computational complexity of processing a single neighborhood is $\mathcal{O}\left(N^3\right)$, with $N^2$ being the size of the neighborhood. However, fast algorithms for calculation of the



Fig. 17. The *U* and *V* chrominances of the JPEG-compressed *Lena* image shown in Figure 16(left). Only very few DCT harmonics survived the aggressive quantization, and the structural information is almost completely lost.
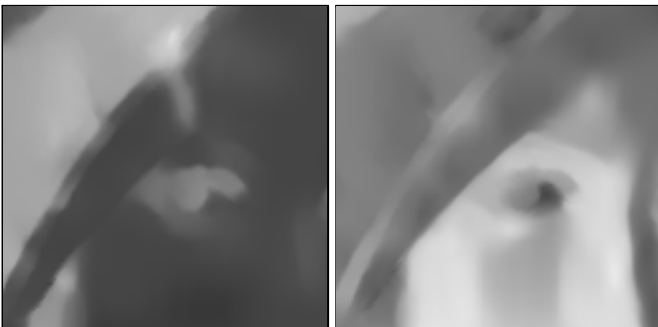


Fig. 18. The chrominances shown in Figure 17 after reconstruction by Pointwise SA-DCT filtering. The blockiness is removed and the structures are faithfully restored.
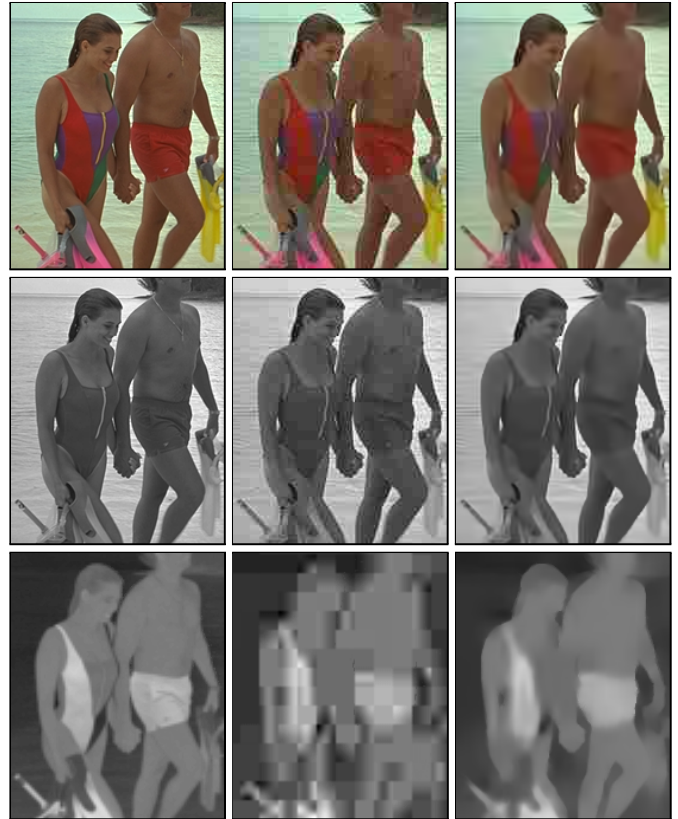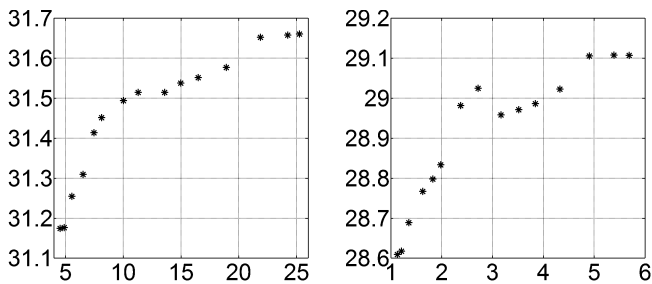
Fig. 20. Denoising performance (PSNR, dB) vs. execution time (seconds) plots for the grayscale $512{\times}512$ *Lena* (left) and $256{\times}256$ *Cameraman* (right) images, $\sigma{=}25$. Execution time refers to the current MATLAB implementation measured on a 1.5 GHz Pentium M CPU.



Fig. 21. Denoising of *Cameraman* ($\sigma{=}25$, observation shown in Figure 7) using B-DCT with fixed block of size $8{\times}8$ (left) and $16{\times}16$ (right). The PSNRs of the two estimates are 28.83dB and 28.63dB, respectively. Compare with the Pointwise SA-DCT estimate shown in Figure 7 (PSNR=29.11dB).

shape-adaptive DCT transforms do exists (e.g. [50]), thus the complexity is $\mathcal{O}\left(N^2 \log N\right)$. In our examples $N$ can vary between 1 and 17, with the most frequent values between 6 and 12. On a more practical level, it is important to remark that highly-optimized hardware platforms (designed for real-time SA-DCT coding of high-resolution video) are available.

Our proposed method is also fully scalable in terms of complexity/performance trade-off. Scalability is achieved by decreasing the number of different scales and especially by limiting the number of overlapping neighborhoods (and thus the overcompleteness). In practice, we do not process a neighborhood $\tilde{U}_x^+$ if $x$ belongs to a sufficient number $M_{\text{overlap}}$ of already-processed neighborhoods: by decreasing $M_{\text{overlap}}$ fewer forward and inverse SA-DCT transforms are computed. Figure 20 shows how a significant acceleration of the algorithm can be achieved in this manner with only a marginal loss in the final estimate's quality. Execution times refer to our current MATLAB implementation of the Pointwise SA-DCT filter (including Anisotropic LPA-ICI and SA-DCT hard-thresholding and Wiener filtering) measured on a 1.5 GHz Pentium M CPU. This MATLAB demonstration software is not optimized in terms of computational efficiency (in particular we do not implement the SA-DCT using fast algorithms), nevertheless its execution time is comparable to other advanced wavelet-based denoising algorithms. As a comparison, on the same machine it takes about 90 seconds to denoise the grayscale *Lena* image ($\sigma{=}25$) using the BLS-GSM algorithm [42] (PSNR=31.69dB) and about 8 seconds using the ProbShrink algorithm [40] (PSNR=31.21dB).

When discussing about complexity/performance trade-off, it is important to underline that in terms of PSNR the empirical Wiener filtering estimate (14) is usually about 0.3-0.5dB better than the simpler hard-thresholding estimate (9) that is used as reference signal for (14). Since the empirical Wiener filtering accounts roughly for half of the overall complexity of the algorithm, the algorithm can be made faster by simply skipping this second-stage filtering. However, it can be seen from the plots in Figure 20 that a much more efficient trade-off is achieved using the above scalability strategy: execution time can be halved (from 25 to 12 seconds for *Lena* and from 5.5 to 2.5 seconds for *Cameraman*) sacrificing as little as 0.15dB in PSNR.

Finally, we wish to note that the impact of the coefficient alignment described in Section IV-C is marginal in terms of

PSNR (about 0.1dB improvement) and completely negligible in terms of complexity (look-up-tables can be utilized), hence its use shall depend exclusively on particular software/hardware design requirements.

### F. Discussion

It is natural to ask to what extent the use of a shape-adaptive transform contributes to the objective and subjective quality achieved by the proposed method. In Figure 21 we show two estimates obtained by the denoising algorithm described in Sections V-B–V-D (hard thresholding and Wiener filtering in transform-domain) using a square block of fixed size ($8{\times}8$ and $16{\times}16$) instead of the adaptively shaped support $\tilde{U}_x^+$. Although in terms of PSNR the results are quite satisfactory, the visual quality of the estimates can be disappointing: several artifacts are visible, including blurring and ringing around the edges. For these two estimates, instead of relying on Equation (7), we used MSE-optimal values of the hard-threshold. We note that the overall visual quality cannot be really improved by increasing or decreasing this threshold, because that would either accentuate the blur or introduce more artifacts. Thus, these may be considered as the best estimates which can be obtained by this algorithm using square blocks of these fixed sizes. The Pointwise SA-DCT estimate shown in Figure 7 not only has higher PSNR, but also exhibits significantly sharper edges and fewer artifacts. While the DCT on blocks of fixed size gives essentially only spatial and frequency selectivity, the Pointwise SA-DCT provides an image representation that combines spatial (adaptation is pointwise), frequency (using SA-DCT harmonics), scale (size of transform basis elements is adaptive), and directional (support is anisotropic) selectivity. The locally-adaptive supports can thus be rightly considered as the main reason of the success demonstrated by the Pointwise SA-DCT filter here and in the aforementioned psychovisual experiment [51], [52]. Further advantages of our method arise when we consider color image processing, since the *structural constraint in luminance-chrominance space* cannot be realized using blocks of fixed size.

### G. Other applications and types of noise

The particular adaptivity of the Pointwise SA-DCT approach can be used for other image restoration applications,

such as image deblurring [13] and inverse-halftoning [8].

While in this paper we considered denoising only from white Gaussian noise, there are no principal limitations to the use of our approach for removal of colored or signal-dependent noise. We refer the reader to [13], where (in the context of deblurring) the Pointwise SA-DCT is used to remove colored noise from regularized deconvolution estimates. Signal-dependent noise can removed exploiting a recursive version Anisotropic LPA-ICI algorithm [11] to identify the adaptive transform supports $\tilde{U}_x^+$. On each support the signal is uniform, hence the noise variance is also uniform and can be well approximated by an adaptive constant $\sigma^2(\tilde{U}_x^+)$ via the variance function [11]. Thus, the SA-DCT filtering and the aggregation can be applied as in Sections V-B–V-D provided that the fixed $\sigma^2$ is replaced by the adaptive $\sigma^2(\tilde{U}_x^+)$.

### H. Software and more examples and results

More simulation results, full-color pictures, and the MAT-LAB software which implements the presented method are available at `http://www.cs.tut.fi/~foi/SA-DCT/` . All the results and figures shown in the present paper can be reproduced using this publicly available software.

## IX. CONCLUSIONS

We presented a novel image filtering approach based on the shape-adaptive DCT transform (SA-DCT). Hard-thresholding and empirical Wiener filtering are performed in SA-DCT domain, with an arbitrarily-shaped transform's support which is adaptively defined for every point in the image. The approach is used for the accurate denoising of grayscale as well as color images. Besides noise removal, the proposed method is also effective in dealing with those artifacts which are often encountered in block-DCT compressed images and videos. Blocking artifacts are suppressed while salient image features are preserved. The luminance-driven shape-adaptive filtering can faithfully reconstruct the missing structural information of the chrominances, thus correcting color-bleeding artifacts. The visual quality of the estimates is high, with sharp detail preservation, clean edges, and without unpleasant artifacts introduced by the fitted transform. The Pointwise SA-DCT algorithms demonstrate a remarkable performance, typically outperforming the best methods known to the authors.

## APPENDIX

For the readers less familiar with the LPA and ICI techniques utilized in Sections III and V-A, in this appendix we give more details on these techniques. For the sake of simplicity, we restrict ourself to the 1D case (thus dropping the subscript $\theta_k$ from notation). This simplification is not essential, because as we described in Section V-A in our implementation we use 1D kernels supported on line segments.

### A. Local Polynomial Approximation (LPA)

The Local Polynomial Approximation (LPA) (e.g. [10]) is a technique which is applied for nonparametric estimation using a polynomial data fit in a sliding window. The polynomial
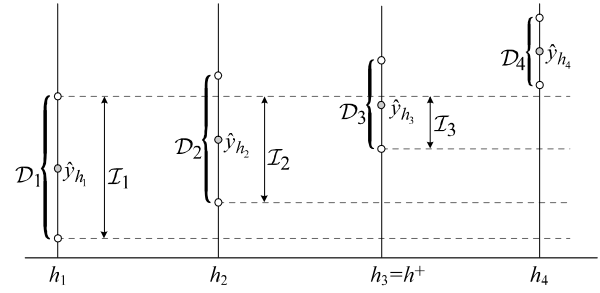


Fig. 22.   The Intersection of Confidence Intervals (ICI) rule.

order $m$ and the window function $w$ characterize the LPA. The LPA estimates are calculated by convolution against a kernel $g = \mathbf{w}\phi\boldsymbol{\Phi}^{-1}[1\ 0\cdots 0]^T$, where $\mathbf{w}= \operatorname{diag} w$ is the diagonal matrix composed by the weights $w$, $\phi$ is a vector of $m+1$ polynomial functions (basis) $\phi_n = \frac{v^n}{n!}$, $n = 0,\ldots,m$, and $\boldsymbol{\Phi}= \phi^T\mathbf{w}\phi$ is the Gramian matrix (formed by the inner products of the basis elements against each other).

Starting from a *basic* window function $w$, one can obtain LPA's of different bandwidths/scales using scaled windows $w_h = w(\cdot/h)$, where $h \in \mathbb{R}^+$ is the *scale* parameter. The corresponding kernels are denoted as $g_h$. It is common practice to use compactly supported window functions. In this case, by using a basic window $w$ of unit length, we obtain that $h$ coincides with the length of the window $w_h$. Hence, window length (size), scale, and bandwidth become interchangeable concepts.

The choice of the scale parameter is crucial when dealing with noisy data, because it controls the amount of smoothing introduced by the local approximation. A large $h$ corresponds to a larger window and therefore to smoother estimates, with lower variance and typically increased estimation bias. A small $h$ corresponds to noisier estimates, less biased, and with higher variance. Thus, the scale parameter $h$ controls the trade-off between bias and variance in the LPA estimates.

### B. Intersection of Confidence Intervals (ICI) rule

The Intersection of Confidence Intervals (ICI) rule [21], [24] is a criterion used for the adaptive selection of the size (length/scale) of the LPA window. Let $x$ be a fixed estimation point/pixel. The LPA estimates $\hat{y}_{h_j}(x) = (z \circledast g_{h_j})(x)$ are calculated for a set $H = \{h_j\}_{j=1}^J$ of increasing scales $h_1 < \cdots < h_J$. The goal of the ICI is to select among these given estimates $\{\hat{y}_{h_j}(x)\}_{j=1}^J$ an adaptive estimate $\hat{y}_{h^+(x)}(x)$, $h^+(x) \in H$, such that $\hat{y}_{h^+(x)}(x)$ is close to an "ideal" estimate $\hat{y}_{h^*(x)}(x)$ which minimizes the MSE with respect to the variation of the scale $h$ (note that $h^*(x)$ does not necessarily belong to $H$). Roughly speaking, the estimate $\hat{y}_{h^+(x)}(x)$ is the "best" among the given ones.

The ICI rule is as follows: *Consider the intersection of confidence intervals* $\mathcal{I}_j = \bigcap_{i=1}^j \mathcal{D}_i$, *where* $\mathcal{D}_i = \left[\hat{y}_{h_i}(x) - \Gamma\sigma_{\hat{y}_{h_i}(x)}, \hat{y}_{h_i}(x) + \Gamma\sigma_{\hat{y}_{h_i}(x)}\right]$, $\sigma_{\hat{y}_{h_i}(x)} = \operatorname{std}\{\hat{y}_{h_i}(x)\}$ *is the standard deviation of* $\hat{y}_{h_i}(x)$, *and* $\Gamma > 0$ *is a threshold parameter. Let* $j^+$ *be the largest of the indexes* $j$ *for which* $\mathcal{I}_j$ *is non-empty,* $\mathcal{I}_{j^+} \neq \varnothing$ *and* $\mathcal{I}_{j^++1} = \varnothing$. *The adaptive scale* $h^+(x)$ *is defined as* $h^+(x) = h_{j^+}$ *and the adaptive estimate is thus* $\hat{y}_{h^+(x)}(x)$.

An illustration of the ICI is given in Figure 22. The standard-deviations of the LPA estimates can be easily calculated from the $\ell^2$-norm of the corresponding kernel as $\sigma_{\hat{y}_{h_j}(x)} = \text{std}\left\{\hat{y}_{h_j}(x)\right\} = \sigma\left\|g_{h_j}\right\|_2$. Since the scales are increasing, the standard-deviations are decreasing and the confidence intervals shrink as $j$ increases. Therefore, in the intersections we are testing estimates with progressively lower variance. The rationale behind the ICI is that the estimation bias is not too large as long as the intersections are non-empty. In practice this means that the ICI adaptively allows the maximum level of smoothing, stopping before oversmoothing begins. Asymptotically, the LPA-ICI adaptive estimator allows to get a near-optimal quality of signal recovery [21].

## REFERENCES

[1] Averbuch, A., A. Schclar, and D.L. Donoho, "Deblocking of block-transform compressed images using weighted sums of symmetrically aligned pixels", *IEEE Trans. Image Process.*, vol. 14, no. 2, pp. 200-212, Feb. 2005.

[2] Acocella, E.C., and A. Alcaim, "Alignment by Phase of Vertical Coefficients in SA-DCT", *IEEE Signal Process. Lett.*, vol. 8, no. 2, pp. 42-44, Feb. 2001.

[3] Bi, M., S.H. Ong, and Y.H. Ang, "Coefficient grouping method for shape-adaptive DCT", *Electron. Lett.*, vol. 32, no. 3, pp. 201-202, Feb. 1996.

[4] Bi, M., S.H. Ong, and Y.H. Ang, "Comment on Shape-Adaptive DCT for generic coding of video", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 6, pp. 686-688, Dec. 1996.

[5] Chen, K.-H., J. Guo, J.-S. Wang, C.-W. Yeh, and J.-W. Chen, "An energy-aware IP core design for the variable-length DCT/IDCT targeting at MPEG4 shape-adaptive transforms", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 5, pp. 704-714, May 2005.

[6] Chen, T., H.R. Wu, and B. Qiu, "Adaptive postfiltering of transform coefficients for the reduction of blocking artifacts", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 5, pp. 584-602, Aug. 2001.

[7] Cho, D., and T.D. Bui, "Multivariate statistical modeling for image denoising using wavelet transforms", *Signal Process.: Image Comm.*, vol. 20, no. 1, pp. 77-89, Jan. 2005.

[8] Dabov, K., A. Foi, V. Katkovnik, and K. Egiazarian, "Inverse halftoning by pointwise shape-adaptive DCT regularized deconvolution", *Proc. 2006 Int. TICSP Workshop Spectral Meth. Multirate Signal Process., SMMSP 2006*, Florence, Sep. 2006.

[9] Donoho, D.L., and I.M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage", *Biometrika*, n. 81, pp. 425-455, 1994.

[10] Fan, J., and I. Gijbels, *Local polynomial modelling and its application*, Chapman and Hall, London, 1996.

[11] Foi, A., R. Bilcu, V. Katkovnik, and K. Egiazarian, "Anisotropic local approximations for pointwise adaptive signal-dependent noise removal", *Proc. XIII Eur. Signal Proc. Conf., EUSIPCO 2005*, Antalya, Sep. 2005.

[12] Foi, A., V. Katkovnik, and K. Egiazarian, "Pointwise shape-adaptive DCT as an overcomplete denoising tool", *Proc. 2005 Int. TICSP Workshop Spectral Meth. Multirate Signal Process., SMMSP 2005*, pp. 164-170, Riga, June 2005.

[13] Foi, A., K. Dabov, V. Katkovnik, and K. Egiazarian, "Shape-adaptive DCT for denoising and image reconstruction", *Proc. SPIE El. Imaging 2006, Image Process.: Algorithms and Systems V*, 6064A-18, Jan. 2006.

[14] Foi, A., V. Katkovnik, and K. Egiazarian, "Pointwise shape-adaptive DCT denoising with structure preservation in luminance-chrominance space", *Proc. 2nd Int. Workshop Video Process. Quality Metrics Consum. Electron., VPQM2006*, Scottsdale, AZ, Jan. 2006.

[15] Foi, A., V. Katkovnik, and K. Egiazarian, "Pointwise shape-adaptive DCT for high-quality deblocking of compressed color images", *Proc. 14th Eur. Signal Process. Conf., EUSIPCO 2006*, Florence, Sep. 2006.

[16] Foi, A., and V. Katkovnik, "From local polynomial approximation to pointwise shape-adaptive transforms: an evolutionary nonparametric regression perspective", *Proc. 2006 Int. TICSP Workshop Spectral Meth. Multirate Signal Process., SMMSP 2006*, Florence, Sep. 2006.

[17] Foi, A., V. Katkovnik, K. Egiazarian, and J. Astola, "A novel anisotropic local polynomial estimator based on directional multiscale optimizations", *Proc. 6th IMA Int. Conf. Math. Signal Process.*, pp. 79-82, Cirencester, 2004.

[18] Foi, A., D. Paliy, V. Katkovnik, and K. Egiazarian, "Anisotropic nonparametric image restoration demobox" (MATLAB software), *Local Approximations in Signal and Image Processing (LASIP) Project*, http://www.cs.tut.fi/~lasip/ , 2005.

[19] Gilge, M., T. Engelhardt, and R. Mehlan, "Coding of arbitrarily shaped image segments based on a generalized orthogonal transform", *Signal Process.: Image Comm.*, vol. 1, no. 2, pp. 153-180, Oct. 1989.

[20] Gilge, M., "Region oriented transform coding (ROTC) of images", *Proc. of Int. Conf. on Acoustics, Speech, and Signal Processing, ICASSP-90*, Albuquerque, USA, vol. 4, pp. 2245-2248, Apr. 1990.

[21] Goldenshluger, A., and A. Nemirovski, "On spatial adaptive estimation of nonparametric regression", *Math. Meth. Statistics*, vol. 6, pp. 135-170, 1997.

[22] Guleryuz, O.G., "Weighted overcomplete denoising", *Proc. Asilomar Conf. Signals Syst. Comput.*, Pacific Grove, CA, Nov. 2003.

[23] Hsung, T.-C., and D.P.-K. Lun, "Application of singularity detection for the deblocking of JPEG decoded images", *IEEE Trans. Circuits Syst. II*, vol. 45, no. 5, pp. 640-644, May 1998.

[24] Katkovnik, V., "A new method for varying adaptive bandwidth selection", *IEEE Trans. Signal Process.*, vol. 47, no. 9, pp. 2567-2571, 1999.

[25] Katkovnik, V., K. Egiazarian, and J. Astola, "Adaptive window size image de-noising based on intersection of confidence intervals (ICI) rule", *J. Math. Imaging and Vision*, vol. 16, no. 3, pp. 223-235, 2002.

[26] Katkovnik, V., A. Foi, K. Egiazarian, and J. Astola, "Directional varying scale approximations for anisotropic signal processing", *Proc. XII Eur. Signal Process. Conf., EUSIPCO 2004*, pp. 101-104, Vienna, Sep. 2004.

[27] Kauff, P., and K. Schuur, "Shape-adaptive DCT with block-based DC separation and $\Delta$DC correction", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 3, pp. 237-242, 1998.

[28] Kaup, A., and S. Panis, "On the performance of the shape adaptive DCT in object-based coding of motion compensated difference images", *Proc. of 1997 Picture Coding Symposium*, pp. 653-657, 1997.

[29] Kervrann, C., and J. Boulanger, "Optimal Spatial Adaptation for Patch-Based Image Denoising", *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 2866-2878, Oct. 2006.

[30] Kinane, A., V. Muresan, and N. O'Connor, "Optimal adder-based hardware architecture for the DCT/SA-DCT", *Proc. SPIE Visual Comm. Image Process. Conf., VCIP 2005*, Beijing, China, 12-15 July 2005.

[31] Kinane, A., A. Casey. V. Muresan, and N. O'Connor, "FPGA-based conformance testing and system prototyping of an MPEG-4 SA-DCT hardware accelerator", *IEEE 2005 Int. Conf. on Field-Programmable Technology, FPT' 05*, Singapore, Dec. 2005.

[32] Koenen, R., "Overview of the MPEG-4 Standard", ISO/IEC JTC1/SC29/WG11 Doc. N3536, July 2000.

[33] Liew, A.W.C., and H. Yan, "Blocking Artifacts Suppression in Block-Coded Images Using Overcomplete Wavelet Representation", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 4, pp. 450-461, Apr. 2004.

[34] Marsi, S., R. Castagno, and G. Ramponi, "A simple algorithm for the reduction of blocking artifacts in images and its implementation", *IEEE Trans. Consum. Electron.*, vol. 44, no. 3, pp. 1062-1070, Aug. 1998.

[35] "MPEG-4 video verification model version 18.0 (VM-18)", ISO/IEC JTC1/SC29/WG11, Doc. N3908, 2001.

[36] MPEG-4: Information technology - Coding of audio-visual objects - Part 2: Visual, ISO/IEC 14496-2:2001, Dec. 2001.

[37] O'Connor, N., S. Sav, T. Adamek, V. Mezaris, I. Kompatsiaris, T.Y. Lu, E. Izquierdo, C. Bennström, and J. Casas, "Region and object segmentation algorithms in the Qimera segmentation platform", *Proc. Third Int. Workshop on Content-Based Multimedia Indexing (CBMI03)*, Rennes, pp. 381-388, 2003.

[38] Ostermann, J., E.S. Jang, J. Shin, and T. Chen, "Coding of arbitrarily shaped video objects in MPEG-4", *Proc. Int. Conf. Image Process., ICIP 1997*, pp. 496-499, 1997.

[39] Paek, H., R.C. Kim, and S.U. Lee, "On the POCS-based postprocessing technique to reduce blocking artifacts in transform coded images," *IEEE Trans. Circuits Sys. Video Technology*, vol. 8, pp. 358-367, June 1998.

[40] Pizurica, A., and W. Philips, "Estimating probability of presence of a signal of interest in multiresolution single- and multiband image

denoising", *IEEE Trans. Image Process.*, vol. 15, no. 3, pp. 654-665, Mar. 2006.

[41] Plataniotis, K.N., and A.N. Venetsanopoulos, *Color image processing and applications*, Springer-Verlag New York, Inc., New York, NY, 2000.

[42] Portilla, J., V. Strela, M. Wainwright, and E.P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain", *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1338-1351, Nov. 2003.

[43] Rosenholts, R., and A. Zakhor, "Iterative procedures for reduction of blocking artifacts in transform domain image coding", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 2, no. 2, pp. 91-95, Mar. 1992.

[44] Scheunders, P., "Wavelet thresholding of multivalued images", *IEEE Trans. Image Process.*, vol. 13, pp. 475-483, Apr. 2004.

[45] Scheunders, P., and J. Driesen, "Least squares interband denoising of color and multispectral images", *Proc. IEEE Int. Conf. Image Process., ICIP 2004*, Singapore, Oct. 2004.

[46] Sendur, L., and I. Selesnick, "Bivariate Shrinkage Functions for Wavelet-Based Denoising Exploiting Interscale Dependency", *IEEE Trans. Signal Process.*, vol. 50, no. 11, pp. 2744-2756, 2002.

[47] Sikora, T., "Low complexity shape-adaptive DCT for coding of arbitrarily shaped image segments", *Signal Process.: Image Comm.*, vol. 7, pp. 381-395, 1995.

[48] Sikora, T., S. Bauer, and B. Makai, "Efficiency of shape-adaptive 2-D transforms for coding of arbitrarily shaped image segments," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, no. 3, pp. 254-258, June 1995.

[49] Sikora, T., and B. Makai, "Shape-adaptive DCT for generic coding of video", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, no. 1, pp. 59-62, 1995.

[50] Stasinski, R., and J. Konrad, "Reduced-complexity shape-adaptive DCT for region-based image coding", *Proc. IEEE Int. Conf. Image Process., ICIP 1998*, Chicago, pp. 114-118, Oct. 1998.

[51] Van der Weken, D., E. Kerre, E. Vansteenkiste, and W. Philips, "Evaluation of fuzzy image quality measures using a multidimensional scaling framework", *Proc. 2nd Int. Workshop Video Process. Quality Metrics Consum. Electron., VPQM2006*, Scottsdale, AZ, Jan. 2006.

[52] Vansteenkiste, E., D. Van der Weken, W. Philips, and E. Kerre, "Perceived image Quality Measurement of state-of-the-art Noise Reduction Schemes", *Lecture Notes in Computer Science 4179 - ACIVS 2006*, pp. 114-124, Springer, Sep. 2006.

[53] Wu, S., H. Yan, and Z. Tan, "An efficient wavelet-based deblocking algorithm for highly compressed images", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 11, pp. 1193-1198, Nov. 2001.

[54] Xiong, Z., M. Orchard, and Y. Zhang, "A deblocking algorithm for JPEG compressed images using overcomplete wavelet representations", *IEEE Trans. Circuits Syst. Video Techn.*, vol. 7, no. 4, pp. 433-437, Apr. 1997.

[55] Yang, Y., N.P. Galatsanos, and A.K. Katsaggelos, "Regularized reconstruction to reduce blocking artifacts of block discrete cosine transform compressed images", *IEEE Trans. Circuits Syst. Video Techn.*, vol. 3, no. 6, pp. 421-432, Dec. 1993.

[56] Yang, Y., N.P. Galatsanos, and A.K. Katsaggelos, "Projection-based spatially adaptive reconstruction of block-transform compressed images", *IEEE Trans. Image Process.*, vol. 4, no. 7, pp. 896-908, July 1995.

**V** ladimir 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 and since 2003 he is a Research Professor with the Institute of Signal Processing, Tampere University of Technology, Tampere, Finland. He has published seven books and more than 200 papers. His research interests include stochastic signal processing, linear and nonlinear filtering, nonparametric estimation, imaging, nonstationary systems, and time-frequency analysis.

**K** aren 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 Transforms and Spectral Methods Group. His research interests are in the areas of applied mathematics, signal processing, and digital logic.

**A** lessandro Foi has received the M.Sc. and the Ph.D. degrees in mathematics from Università degli Studi di Milano (Italy) in 2001 and from Politecnico di Milano in 2005, respectively. His research interests include mathematical and statistical methods for signal processing, functional analysis, and harmonic analysis. Currently, he is a researcher at the Institute of Signal Processing, Tampere University of Technology (Finland). His work focuses on spatially adaptive algorithms for denoising and deblurring of digital images and on noise modeling for digital imaging sensors.