A NONLOCAL AND SHAPE-ADAPTIVE TRANSFORM-DOMAIN COLLABORATIVE FILTERING

Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian

Department of Signal Processing, Tampere University of Technology
P.O. Box 553, 33101, Tampere, Finland
email: firstname.lastname@tut.fi

ABSTRACT

We propose an image denoising method that exploits both nonlocal image modeling and locally adaptive anisotropic estimation. The method uses grouping of adaptive-shape neighborhoods whose surrounding square supersets have been found similar by a block-matching procedure. The data defined on these grouped neighborhoods is stacked together, resulting in 3-D data structures which are generalized cylinders with adaptive-shape cross sections. Because of the similarity, which follows from the matching, and because of the adaptive selection of the shape of the neighborhoods, these 3-D groups are characterized by a high correlation along all the three dimensions. We apply a 3-D decorrelating transform, computed as a separable composition of the Shape-Adaptive DCT (SA-DCT) and a 1-D orthonormal transform, and subsequently attenuate the noise by spectrum shrinkage with hard-thresholding or Wiener filtering. Inversion of the 3-D transform produces individual estimates for all grouped neighborhoods. These estimates are returned to their original locations and aggregated with other estimates coming from different groups.

Overall, this method generalizes two existing filters: the BM3D filter, which uses grouping of fixed-size square blocks, and the Pointwise SA-DCT filter, which exploits shrinkage on adaptive-shape supports. We show that the developed method inherits the strengths of both filters, resulting in a very effective and flexible tool for image denoising.

1. INTRODUCTION

Image denoising is a fundamental problem in image processing and a lot of research has been dedicated it. Some of the recent and most successful advances in the field are the methods [15, 7, 3, 5, 1, 13, 2, 14]. Among these, are two methods by the present authors, the BM3D filter [2] and the Pointwise SA-DCT filter (P.SA-DCT) [5], which are used as a basis for this work.

The BM3D filter exploits a specific nonlocal image modeling [9] through a procedure termed *grouping and collaborative filtering*. Grouping finds mutually similar 2-D image blocks and stacks them together in 3-D arrays. Collaborative filtering produces individual estimates of all grouped blocks by filtering them jointly, through transform-domain shrinkage of the 3-D arrays (groups). In doing so, BM3D relies on two strong characteristics of natural images. First, the abundance of mutually similar patches and, second, that the content of small blocks is locally highly correlated. On these assumptions, the group enjoys correlation in all three dimensions and a sparse representation of the true signal is obtained by applying a decorrelating 3-D transform on the group. The subsequent shrinkage achieves effective noise attenuation thanks to sparsity. We have shown [2] that the BM3D filter is a very efficient and powerful denoiser. Its results are still beyond the capabilities of most of the more recent and advanced algorithms. Of course, BM3D is particularly successful when plenty of matching blocks can be found (e.g., textures, regular shaped image structures, or uniform areas), i.e. where the nonlocal modeling is suitable. On the other hand, the assumption that image content is highly correlated on a square block of fixed size is sometimes very artificial. Blocks containing small image details, singularities, or sharp and curved edges are examples where a non-adaptive transform is not able to deliver a sparse representation. Thus, for these blocks, the BM3D filter may introduce certain artifacts and the denoising is not very effective. Unfortunately, these are often the very parts of the image where the visual attention is mainly focused.

The Pointwise SA-DCT filter [5] works differently. It exploits *pointwise-adaptive arbitrarily shaped neighborhoods*, adaptive to image content in such a way that the signal contained in a neighborhood is highly homogeneous. The subsequent application of the (2-D) shape-adaptive discrete cosine transform (SA-DCT) [12] on such a neighborhood results in a sparse representation of the true signal, which enables effective noise attenuation by transform shrinkage. Due to the adaptivity of neighborhoods to local image details, the P.SA-DCT filter achieves exceptional preservation of edges and singularities, as shown [5, 4]. However, a drawback of the nonparametric local-homogeneous image model, assumed by the Pointwise SA-DCT filter, becomes evident for texture-rich regions. There, local homogeneity is very limited and, thus, the performance of the filter is not satisfactory [4]. Moreover, the P.SA-DCT is a *local* filter, and as such it cannot take advantage of the abundance of repeated structures and patterns found in natural images.

From the mentioned properties of the BM3D and the P.SA-DCT filters, one can infer that their strengths and weaknesses are complementing each other. This is the motivation that leads us to propose in this work an image denoising method that simultaneously generalizes these two filters. The new method performs *grouping* of mutually similar *adaptive-shape neighborhoods*. Thus, the 3-D groups become generalized cylinders with adaptive-shape cross sections (as opposed to the groups in BM3D, which are square prisms with fixed-size square cross sections). In this manner, both the nonlocal image modeling and the local adaptivity to image features are exploited. In particular, the spatial correlation within the 2-D neighborhoods is increased as compared with the BM3D. After the grouping, we apply a 3-D transform that is a separable composition of the SA-DCT and a 1-D orthonormal transform. Noise is then attenuated by spectrum shrinkage with either hard-thresholding or Wiener filtering. Similar to the BM3D, inversion of the 3-D transform produces individual estimates for all grouped neighborhoods. These estimates are returned to their original locations and aggregated with estimates coming from other groups.

We show that this developed method, denominated SA-BM3D, inherits the strengths of both the P.SA-DCT and the BM3D filters and at the same time overcomes their weaknesses.

2. PRELIMINARIES

We consider noisy observations z of the form

$$z(x) = y(x) + \eta(x), \quad x \in X, \tag{1}$$

where $y : X \to \mathbb{R}$ is the original grayscale image, η is i.i.d. Gaussian white noise, $\eta(\cdot) \sim \mathcal{N}(0, \sigma^2)$, and x is a spatial variable belonging

This work was supported by the Academy of Finland (application no. 213462, Finnish Programme for Centres of Excellence in Research 2006-2011, and application no. 118312, Finland Distinguished Professor Programme 2007-2010); the first author was supported by TISE graduate school.

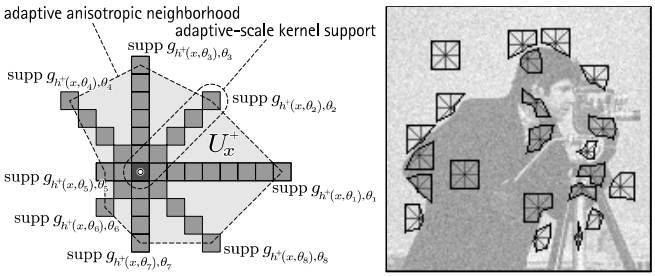

Figure 1: Fast implementation of the LPA-ICI anisotropic neighborhoods. "Linewise" one-dimensional directional LPA kernels are used for 8 directions. The anisotropic neighborhood U_x^+ is constructed as the polygonal hull of the adaptive-scale kernels' supports (left). Thus, only the adaptive scales 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).

to the image domain $X \subset \mathbb{Z}^2$.

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 elements) of U. The symbol "$\circledast$" stands for the convolution operation.

2.1 Adaptive anisotropic neighborhoods

As in [5], we use the local polynomial approximation-intersection of confidence intervals (LPA-ICI) technique [11] in order to associate, to each $x \in X$, and adaptive-shape neighborhood in which the image is homogenous.

2.1.1 Directional pointwise adaptive scales by LPA-ICI

For each of the eight directions $\theta_k = \frac{(k-1)}{4}\pi,\ k = 1, \ldots, 8$, a varying-scale family of narrow "linewise" directional-LPA [10] 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 [6, 8, 11], and as a result an adaptive scale $h^+(x,\theta_k) \in H$ is defined for every $x \in X$ and for every direction θ_k.

2.1.2 Adaptive-shape neighborhood

The anisotropic 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 1. We note that, in our 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 θ_k (i.e. with the length of the support of the corresponding directional kernel). Thus, in order to construct any neighborhood U_x^+ it suffices to know only the adaptive scales $\left\{h^+(x,\theta_k)\right\}_{k=1}^{8}$ for all $x \in X$.

Let us remind that being convolution kernels, the LPA kernels g_{h,θ_k} are always "centered" at the origin, therefore U_x^+ is always a neighborhood of the origin. The actual adaptive neighborhood of x, which contains the observations that are used for estimation, is instead

$$\tilde{U}_x^+ = \left\{v \in X : (x - v) \in U_x^+\right\},$$

in other words $\tilde{U}_x^+$ (with tilde) is obtained by translation and mirroring of U_x^+ (without tilde). 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. We remark that neighborhoods $\tilde{U}_{x'}^+, \tilde{U}_{x''}^+$ corresponding to adjacent or nearby points x', x'' do usually overlap unless an edge or sharp transition exists between the two points.

Additionally, in this work, we also use the following more general mirrored translates of U_x^+:

$$\begin{aligned}\tilde{U}_{x,x_R}^+ &= \left\{v \in X : (x - v) \in U_{x_R}^+\right\} = \\ &= \left\{v \in X : (x_R - x + v) \in \tilde{U}_{x_R}^+\right\}.\end{aligned}$$

$\tilde{U}_{x,x_R}^+$ is an adaptive neighborhood of x which differs from $\tilde{U}_x^+$ in that $\tilde{U}_{x,x_R}^+$ uses the adaptive scales corresponding to the point x_R and not those corresponding to x itself. Obviously, $\tilde{U}_{x,x}^+ = \tilde{U}_x^+$.

2.2 Shape-Adaptive DCT transform

As in [5], we consider the orthonormal SA-DCT with DC-separation [12]. It means that before applying the orthonormal SA-DCT, we first subtract the data of its mean, which constitutes an extra coefficient that will be processed independently from the actual SA-DCT coefficients. Once the DC is separated, the orthonormal SA-DCT is applied on zero-mean data. The SA-DCT is computed by cascaded application of one-dimensional varying-length orthonormal DCT transforms first on the columns and then on the rows (or vice versa[1]) that constitute the considered region. This process is illustrated in Figure 2. We denote by T_U^{SA} the orthonormal SA-DCT transform corresponding to a region $U \subset X$ and by $V_U \subset \mathbb{Z}^2$ the domain of the T_U^{SA} transform coefficients. Let $T_U^{\mathsf{SA}-1}$ be the inverse transform of T_U^{SA}. The mean of z on U is denoted as $m_U(z) = \frac{1}{|U|}\sum_{x \in U} z(x)$.

We refer the interested reader to [5] for more subtle details about the selection between column/row or row/column processing and the particular coefficient alignment strategies used within the SA-DCT.

3. SA-BM3D ALGORITHM

Similar to the BM3D, the proposed SA-BM3D algorithm (illustrated in Figure 3) exploits a two-step approach where

- in the first step, block-matching is performed on the noisy image and the noise is attenuated by collaborative hard-thresholding,
- in the second step, block-matching is performed on the initial (basic) estimate obtained in the first step, and the noise is attenuated by collaborative empirical Wiener filtering.

Each of the two steps is presented in the following subsections.

3.1 SA-BM3D with hard thresholding (Step 1)

Once the LPA-ICI adaptive scales $\left\{h^+(x,\theta_k)\right\}_{k=1}^{8}$, and thus the adaptive-shape neighborhoods U_x^+, have been found (Section 2.1), for each $x \in X$, the following operations are performed:

- shape-adaptive grouping;
- collaborative hard-thresholding;
- aggregation.

These operations are explained in detail in the following subsections; therein we fix the currently processed coordinate as $x_R \in X$ and denominate it *reference point*.

3.1.1 Shape-adaptive grouping via block-matching

The adaptive neighborhoods $\tilde{U}_x^+$ can be too small for reliable patch-matching, especially when filtering tiny image details in heavy noise. Therefore, the matching for $\tilde{U}_x^+$ needs to be carried out for

[1] Note that even though the SA-DCT is implemented like a separable 2-D transform (using cascaded 1-D transforms on columns and rows), in general it is not separable and different sets of transform coefficients can be obtained when processing the considered region either first column-wise and then row-wise or first row-wise and then column-wise.

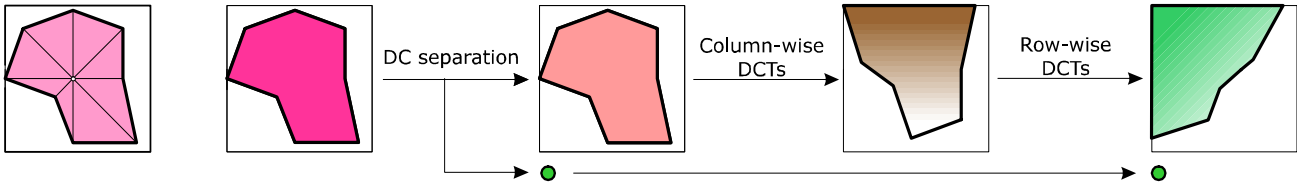

Figure 2: Illustration of the forward SA-DCT with DC-separation.

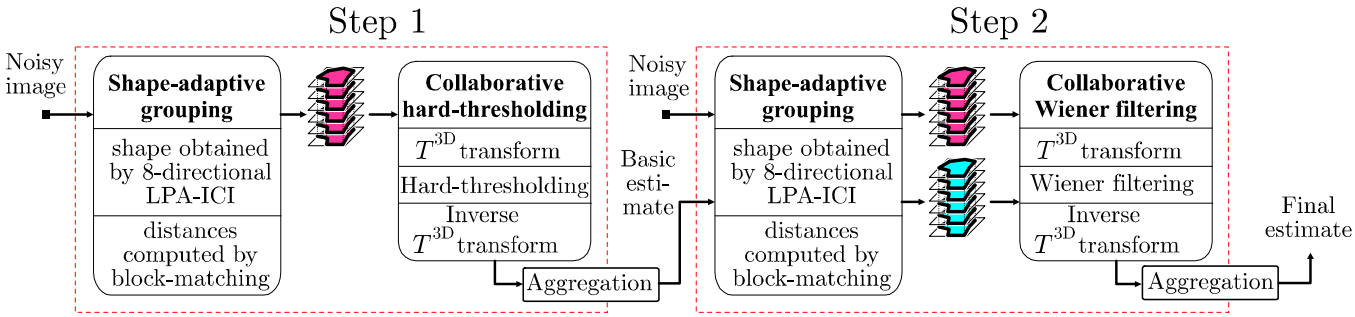

Figure 3: Flowchart of the proposed SA-BM3D image denoising method. Operations surrounded by dashed lines are repeated for each processed coordinate of the input image.

a superset. In particular, we use square blocks as supersets and perform a block-matching procedure like the one in the BM3D algorithm [2]. In what follows, we define a mapping that associates to every reference point $x_R \in X$ a block that can be used as the reference one in the matching needed for the grouping of $\tilde{U}^+_{x_R}$.

Mapping. Let $x \in X$ and denote by $\tilde{B}_x \subset \mathbb{Z}^2$ be the square block of size $(2h_{\text{max}}-1)\times(2h_{\text{max}}-1)$ centered at x, where $h_{\text{max}} = \max\{H\}$. Let $\mathbb{B}$ be the collection of all such blocks which are entirely contained in X, $\mathbb{B} = \{\tilde{B}_x : x \in X,\ \tilde{B}_x \subset X\}$. Note that for $h_{\text{max}} > 1$ the cardinality of $\mathbb{B}$ is strictly smaller than $|X|$ because if x is close enough to the boundary ∂X of X, then $\tilde{B}_x$ would cross ∂X and hence $\tilde{B}_x \not\subseteq X$ and $\tilde{B}_x \notin \mathbb{B}$. We indicate by $X_{\mathbb{B}} \subset X$ the set of points for which we can construct a block belonging to $\mathbb{B}$, $X_{\mathbb{B}} = \{x \in X : \tilde{B}_x \in \mathbb{B}\}$. To every $x \in X$ we can associate a point $x_{\mathbb{B}} \in X_{\mathbb{B}}$ such that the magnitude $\|\delta_{\mathbb{B}}(x)\|_2$ of $\delta_{\mathbb{B}}(x) = x_{\mathbb{B}} - x$ is minimal. For a rectangular X, because of convexity of $X_{\mathbb{B}}$, the surjective mapping $x \mapsto x_{\mathbb{B}}$ is univocally defined and so is $\delta_{\mathbb{B}}(x)$. Note that $\delta_{\mathbb{B}}(x) \neq 0$ only for x sufficiently close to the boundary ∂X of X.

Block-matching. For each point $x \in X_{\mathbb{B}}$, we produce grouping by block-matching within the image z. That is, for each block $\tilde{B}_x$ we look for "similar" blocks $\tilde{B}_{x'}$ whose range distance $d_z(x,x')$ with respect to $\tilde{B}_x$,

$$d_z(x,x') = \left\| z_{|\tilde{B}_x} - z_{|\tilde{B}_{x'}} \right\|_2,$$

is smaller than a fixed threshold $\tau^{\text{ht}}_{\text{match}} \geq 0$. Thus, we construct the set S_x so that it contains the central points of the found blocks:

$$S_x = \left\{ x' \in X_{\mathbb{B}} : d_z(x,x') \leq \tau^{\text{ht}}_{\text{match}} \right\}. \tag{2}$$

The threshold $\tau^{\text{ht}}_{\text{match}}$ is the maximum d_z-distance for which two blocks are considered similar[2]. Obviously $d_z(x,x) = 0$, which implies that the cardinality $|S_x| \geq 1$ for any $x \in X_{\mathbb{B}}$.

Shape-adaptive grouping. Let now $x_R \in X$ be a reference point and define $x^{\delta}_R = x_R + \delta_{\mathbb{B}}(x_R)$. Using the result $S_{x^{\delta}_R}$ from the block-matching, we associate to the reference point x_R not only its own adaptive neighborhood $\tilde{U}^+_{x_R}$, but a whole collection (disjoint union) $\widetilde{\mathbb{U}}_{x_R}$ of neighborhoods having the same shape and defined as[3]

$$\widetilde{\mathbb{U}}_{x_R} = \coprod_{x+\delta_{\mathbb{B}}(x_R) \in S_{x^{\delta}_R}} \tilde{U}^+_{x,x_R}. \tag{3}$$

Let us observe that all neighborhoods in $\widetilde{\mathbb{U}}_{x_R}$ have the same shape, which is completely determined by the adaptive scales $\{h^+(x_R,\theta_k)\}_{k=1}^{8}$ at x_R. This also implies that for all $\tilde{U}^+_{x,x_R} \in \widetilde{\mathbb{U}}_{x_R}$ the domain of the corresponding SA-DCT coefficients $V_{\tilde{U}^+_{x,x_R}}$ is one and the same and coincides with $V_{\tilde{U}^+_{x_R}}$.

At the current reference point x_R, a group is built by stacking together the noisy patches $z_{|\tilde{U}^+_{x,x_R}}$, $\tilde{U}^+_{x,x_R} \in \widetilde{\mathbb{U}}_{x_R}$. This group is a 3-D data array defined on the generalized cylinder $\tilde{U}^+_{x_R} \times \left\{1,\ldots,\left|S_{x^{\delta}_R}\right|\right\}$, as illustrated in Figure 4. In compact form, the group is denoted as $\mathbf{Z}_{x_R} : \widetilde{\mathbb{U}}_{x_R} \to \mathbb{R}$.

3.1.2 *Collaborative hard-thresholding*

Given a group $\mathbf{Z}_{x_R}$, collaborative filtering is realized as shrinkage in a 3-D transform domain. Here, the 3-D transform T^{3D} is a composition of the (2-D) SA-DCT with DC-separation on each neighborhood of $\widetilde{\mathbb{U}}_{x_R}$ with an orthonormal 1-D transform T^{1D} applied along the third dimension of the group. As in the BM3D algorithm [2], we require T^{1D} to have a DC term. The three steps of the collaborative hard-thresholding are as follows.

1. Forward T^{3D} transform (illustrated in Figure 4).
   - For each $\tilde{U}^+_{x,x_R} \in \widetilde{\mathbb{U}}_{x_R}$:
     - compute mean value $m_{\tilde{U}^+_{x,x_R}}(z) = \frac{1}{\left|\tilde{U}^+_{x,x_R}\right|} \sum_{v \in \tilde{U}^+_{x,x_R}} z(v)$;

[2] The influence of noise on d_z is studied in [2]. In case of heavy noise, to reduce its influence on d_z, we can exploit a coarse prefiltering embedded within d_z, as described in [2].

[3] The set (3) must not be interpreted as the mere union of the neighborhoods $\tilde{U}^+_{x,x_R}$ such that $x+\delta_{\mathbb{B}}(x_R) \in S_{x^{\delta}_R}$. While, for simplicity, we may write $\widetilde{\mathbb{U}}_{x_R} = \{\tilde{U}^+_{x,x_R} : x+\delta_{\mathbb{B}}(x_R) \in S_{x^{\delta}_R}\}$, a more proper notation is actually

$$\widetilde{\mathbb{U}}_{x_R} = \left\{ \left(\tilde{U}^+_{x,x_R}, x\right) : x+\delta_{\mathbb{B}}(x_R) \in S_{x^{\delta}_R} \right\} \subset X \times X,$$

because we need to distinguish between different neighborhoods coming from different elements of $S_{x^{\delta}_R}$.

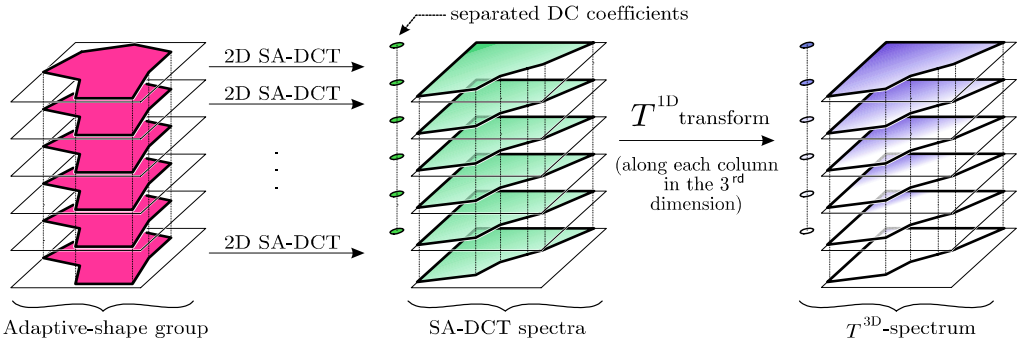

Figure 4: Illustration of applying T^{3D} on the adaptive-shape group $\mathbf{Z}_{x_R}$.

– compute the SA-DCT spectrum $\varphi_{z,x,x_R} : V_{\tilde{U}^+_{x_R}} \to \mathbb{R}$ as

$$T^{SA}_{\tilde{U}^+_{x,x_R}}\left(z_{|\tilde{U}^+_{x,x_R}} - m_{\tilde{U}^+_{x,x_R}}(z)\right);$$

– define DC coefficient $\varphi^{DC}_{z,x,x_R} = \left|\tilde{U}^+_{x_R}\right|^{1/2} m_{\tilde{U}^+_{x,x_R}}(z)$.

At this stage, for each $\tilde{U}^+_{x,x_R} \in \widetilde{\mathbb{U}}_{x_R}$, we have obtained one DC coefficient φ^{DC}_{z,x,x_R} and $\left|\tilde{U}^+_{x_R}\right|$ SA-DCT coefficients φ_{z,x,x_R}. Thus, we have a total of $\left(1+\left|\tilde{U}^+_{x_R}\right|\right)\left|S_{x^\delta_R}\right|$ coefficients, structured as a 3-D array $\left(\{\mu\} \amalg V_{\tilde{U}^+_{x_R}}\right) \times \left\{1,\dots,\left|S_{x^\delta_R}\right|\right\}$, where $\{\mu\}$ is a singleton placeholder for separated DC coefficients φ^{DC}_{z,x,x_R}.

- Apply T^{1D} along the vector $\varphi^{DC}_{z,x,x_R} : \{\mu\} \times \left\{1,\dots,\left|S_{x^\delta_R}\right|\right\} \to \mathbb{R}$ and along the vectors $\varphi_{z,x,x_R} : \{\nu\} \times \left\{1,\dots,\left|S_{x^\delta_R}\right|\right\}$, $\nu \in V_{\tilde{U}^+_{x_R}}$.

This second stage provides us with the full 3-D spectrum $T^{3D}\left(\mathbf{Z}_{x_R}\right)$ of the group $\mathbf{Z}_{x_R}$. We emphasize again that there are $\left(1+\left|\tilde{U}^+_{x_R}\right|\right)\left|S_{x^\delta_R}\right|$ spectral coefficients.

2. Shrinkage by hard-thresholding

As in [2], we perform hard-thresholding of the spectrum $T^{3D}\left(\mathbf{Z}_{\widetilde{\mathbb{U}}_{x_R}}\right)$ using the threshold $\gamma_{\text{thr}}\sigma$, where $\gamma_{\text{thr}} > 0$ is a fixed constant. We threshold all coefficients except the DC coefficient in the T^{3D}-spectrum (i.e. the DC coefficient of the T^{1D}-spectrum of the vector φ^{DC}_{z,x,x_R}), which systematically is always preserved. After thresholding, we have $N^{har}_{x_R} \geq 1$ non-thresholded coefficients ("number of harmonics").

3. Inverse T^{3D} transform

Inversion of the T^{3D} transform is computed as follows.

- Apply the inverse of T^{1D} and obtain filtered vectors $\hat{\varphi}^{DC}_{y,x,x_R}$ and $\hat{\varphi}_{y,x,x_R}$ which are estimates of the DCs and SA-DCT spectra of the hypothetical group $\mathbf{Y}_{x_R}$ formed by stacking together patches $y_{|\tilde{U}^+_{x,x_R}}$, $\tilde{U}^+_{x,x_R} \in \widetilde{\mathbb{U}}_{x_R}$, from the unknown noise-free image y.
- For each $\tilde{U}^+_{x,x_R} \in \widetilde{\mathbb{U}}_{x_R}$:
  – define the estimate of the mean $m_{\tilde{U}^+_{x,x_R}}(y)$ as

$$\widehat{m_{\tilde{U}^+_{x,x_R}}(y)} = \left|\tilde{U}^+_{x_R}\right|^{-1/2} \hat{\varphi}^{DC}_{y,x,x_R};$$

  – compute a local estimate $\hat{y}_{x,x_R} : \tilde{U}^+_{x,x_R} \to \mathbb{R}$ of $y_{|\tilde{U}^+_{x,x_R}}$ by inverse SA-DCT of $\hat{\varphi}_{y,x,x_R}$ followed by addition of the mean, $\hat{y}_{x,x_R} = T^{SA\,-1}_{\tilde{U}^+_{x,x_R}}\left(\hat{\varphi}_{y,x,x_R}\right) + \widehat{m_{\tilde{U}^+_{x,x_R}}(y)}$.

In this way, we obtain local estimates for each of the neighborhoods in $\widetilde{\mathbb{U}}_{x_R}$. All these estimates can be denoted in compact group form as $\widehat{\mathbf{Y}}_{x_R} : \widetilde{\mathbb{U}}_{x_R} \to \mathbb{R}$.

3.1.3 *Aggregation*

After performing grouping and collaborative filtering for all $x_R \in X$, we have obtained for each x_R a group $\widehat{\mathbf{Y}}_{x_R}$ of $\left|S_{x^\delta_R}\right|$ of distinct *local* estimates (i.e. filtered data on the adaptive-shape neighborhoods) of y. Overall, we end up with $\sum_{x_R \in X}\left|S_{x^\delta_R}\right|$ local estimates. We note that each local estimate is supported in one of the $|X|$ adaptive neighborhoods $\tilde{U}^+_x$, for $x \in X$. However, provided $x_R \neq x'_R$, even when two neighborhoods $\tilde{U}^+_{x,x_R} = \tilde{U}^+_{x,x'_R}$ coincide, the respective estimates $\hat{y}_{x,x_R}$ and $\hat{y}_{x,x'_R}$ can be different, as they are obtained from filtering possibly different groups $\mathbf{Z}_{x_R}$ and $\mathbf{Z}_{x'_R}$. Thus, the collection of local estimates $\coprod_{x_R \in X}\widehat{\mathbf{Y}}_{x_R} = \coprod_{x_R \in X,\, x+\delta_{\mathbb{B}}(x_R) \in S_{x^\delta_R}} \hat{y}_{x,x_R}$ is a highly redundant and rich representation of the original image y.

In order to obtain a single *global* estimate $\hat{y}^{ht} : X \to \mathbb{R}$ defined on the whole image domain, all these local estimates are averaged together using adaptive weights $w_{x_R} > 0$ in the following convex combination:

$$\hat{y}^{ht} = \frac{\sum_{x_R \in X}\sum_{x+\delta_{\mathbb{B}}(x_R) \in S_{x^\delta_R}} w_{x_R}\hat{y}_{x,x_R}}{\sum_{x_R \in X}\sum_{x+\delta_{\mathbb{B}}(x_R) \in S_{x^\delta_R}} w_{x_R}\chi_{\tilde{U}^+_{x,x_R}}}^{|X}, \tag{4}$$

where the weights w_{x_R} are defined by

$$w_{x_R} = \frac{\sigma^{-2}}{N^{har}_{x_R}\,|\tilde{U}^+_{x_R}|}. \tag{5}$$

3.2 SA-BM3D with Wiener filtering (Step 2)

As in [2, 5], the denoising performance is significantly improved by applying a second step with empirical Wiener filtering. The initial estimate $\hat{y}^{ht}$ obtained in the first step is used for two purposes; first, the shape-adaptive grouping exploits block-matching performed on $\hat{y}^{ht}$ rather than on the noisy image and, second, the magnitude of the spectrum of $\hat{y}^{ht}$ is used to perform empirical Wiener filtering (rather than hard-thresholding as in the first step). For a reference point $x_R \in X$, the shape-adaptive grouping and the collaborative Wiener filtering are given in the following two subsections.

3.2.1 *Shape-adaptive grouping*

Because the noise in $\hat{y}^{ht}$ is assumed to be significantly attenuated, we can obtain more accurate block-matching by replacing the distance d_z by a distance $d_{\hat{y}^{ht}}$ where the similarity between blocks is evaluated on $\hat{y}^{ht}$ instead of z:

$$d_{\hat{y}^{ht}}(x,x') = \left\|\hat{y}^{ht}_{|\tilde{B}_x} - \hat{y}^{ht}_{|\tilde{B}_{x'}}\right\|_2.$$

Thus, we redefine the sets S_x (which contain the coordinates of the matched blocks) as

$$S_x = \left\{ x' \in X_{\mathbb{B}} : d_{\hat{y}^{\text{ht}}}\left(x, x'\right) \leq \tau_{\text{match}}^{\text{wie}} \right\}, \tag{6}$$

where $\tau_{\text{match}}^{\text{wie}} > 0$ is another threshold for the range distance between blocks.

Using (6), we construct the collection of neighborhoods $\widetilde{\mathbb{U}}_{x_R}$ as in (3). Subsequently, $\widetilde{\mathbb{U}}_{x_R}$ is used to build two groups, first, the group $\mathbf{Z}_{x_R} : \widetilde{\mathbb{U}}_{x_R} \to \mathbb{R}$ by stacking together the noisy patches $z_{|\tilde{U}^+_{x,x_R}}$, $\tilde{U}^+_{x,x_R} \in \widetilde{\mathbb{U}}_{x_R}$, and second, $\widehat{\mathbf{Y}}^{\text{ht}}_{x_R} : \widetilde{\mathbb{U}}_{x_R} \to \mathbb{R}$ by stacking together the initial estimate patches $\hat{y}^{\text{ht}}_{|\tilde{U}^+_{x,x_R}}$, $\tilde{U}^+_{x,x_R} \in \widetilde{\mathbb{U}}_{x_R}$. Both groups are defined on the same domain $\widetilde{\mathbb{U}}_{x_R}$.

3.2.2 Collaborative Wiener filtering

The shrinkage of the 3-D spectrum of the group $\mathbf{Z}_{x_R}$ is performed by empirical Wiener filtering which uses the magnitude of the $T^{3\text{D}}$ spectrum of the group $\widehat{\mathbf{Y}}^{\text{ht}}_{x_R}$. Because of the DC-separation, some care is required when these 3-D spectra are defined.

In particular, while the 3-D spectrum of $\mathbf{Z}_{x_R}$ is defined as in Section 3.1.2, the spectrum of $\widehat{\mathbf{Y}}^{\text{ht}}_{x_R}$ is computed by subtracting from $\hat{y}^{\text{ht}}_{|\tilde{U}^+_{x,x_R}}$ the mean $m_{\tilde{U}^+_{x,x_R}}(z)$, $\tilde{U}^+_{x,x_R} \in \widetilde{\mathbb{U}}_{x_R}$. Additionally, the shrinkage coefficients for the DC-vector are computed using directly the corresponding DC-vector with the means of $\hat{y}^{\text{ht}}_{|\tilde{U}^+_{x,x_R}}$. Let us describe precisely these operations.

1. Forward $T^{3\text{D}}$ transform
   - For each $\tilde{U}^+_{x,x_R} \in \widetilde{\mathbb{U}}_{x_R}$:
     - compute mean value $m_{\tilde{U}^+_{x,x_R}}(z) = \frac{1}{\left|\tilde{U}^+_{x,x_R}\right|} \sum_{v \in \tilde{U}^+_{x,x_R}} z(v)$;
     - compute mean value $m_{\tilde{U}^+_{x,x_R}}(\hat{y}^{\text{ht}}) = \frac{1}{\left|\tilde{U}^+_{x,x_R}\right|} \sum_{v \in \tilde{U}^+_{x,x_R}} \hat{y}^{\text{ht}}(v)$;
     - compute the SA-DCT spectrum $\varphi_{z,x,x_R} : V_{\tilde{U}^+_{x_R}} \to \mathbb{R}$ as
     $$T^{\text{SA}}_{\tilde{U}^+_{x,x_R}}\left(z_{|\tilde{U}^+_{x,x_R}} - m_{\tilde{U}^+_{x,x_R}}(z) \right);$$
     - compute the SA-DCT spectrum $\varphi_{\hat{y}^{\text{ht}},x,x_R} : V_{\tilde{U}^+_{x_R}} \to \mathbb{R}$ as
     $$T^{\text{SA}}_{\tilde{U}^+_{x,x_R}}\left(\hat{y}^{\text{ht}}_{|\tilde{U}^+_{x,x_R}} - m_{\tilde{U}^+_{x,x_R}}(z) \right);$$
     (note that we subtract $m_{\tilde{U}^+_{x,x_R}}(z)$ from both $z_{|\tilde{U}^+_{x,x_R}}$ and $\hat{y}^{\text{ht}}_{|\tilde{U}^+_{x,x_R}}$)
     - define DC coefficient $\varphi^{\text{DC}}_{z,x,x_R} = \left|\tilde{U}^+_{x_R}\right|^{1/2} m_{\tilde{U}^+_{x,x_R}}(z)$.
     - define DC coefficient $\varphi^{\text{DC}}_{\hat{y}^{\text{ht}},x,x_R} = \left|\tilde{U}^+_{x_R}\right|^{1/2} m_{\tilde{U}^+_{x,x_R}}\left(\hat{y}^{\text{ht}}\right)$.
   - Like in Section 3.1.2, apply $T^{1\text{D}}$ along the vectors $\varphi^{\text{DC}}_{z,x,x_R}$, $\varphi^{\text{DC}}_{\hat{y}^{\text{ht}},x,x_R}$, φ_{z,x,x_R} and $\varphi_{\hat{y}^{\text{ht}},x,x_R}$.

   This second stage provides us with the full 3-D spectrum $T^{3\text{D}}\left(\mathbf{Z}_{x_R}\right)$ of $\mathbf{Z}_{x_R}$ and with a special reference 3-D spectrum $\check{T}^{3\text{D}}\left(\widehat{\mathbf{Y}}^{\text{ht}}_{x_R}\right)$ of $\widehat{\mathbf{Y}}^{\text{ht}}_{x_R}$. We need to distinguish between $T^{3\text{D}}$ and $\check{T}^{3\text{D}}$ because from $\hat{y}^{\text{ht}}_{|\tilde{U}^+_{x,x_R}}$ we subtract $m_{\tilde{U}^+_{x,x_R}}(z)$ instead of the mean $m_{\tilde{U}^+_{x,x_R}}\left(\hat{y}^{\text{ht}}\right)$. The asymmetry of the DC-separation in the definition of the 3-D transforms for the groups $\mathbf{Z}_{x_R}$ and $\widehat{\mathbf{Y}}_{x_R}$ is to circumvent the intrinsic redundancy in the SA-DCT with DC-separation, which has as spectrum of $\left|\tilde{U}^+_{x_R}\right| + 1$ elements when $\left|\tilde{U}^+_{x_R}\right|$ would suffice. Essentially, by subtracting the same means from both groups, we treat these subtracted means as deterministic terms that would then be inactive in the Wiener filtering of the other coefficients.

2. Wiener filtering

   We obtain an estimate $\widehat{T^{3\text{D}}\left(\mathbf{Y}_{x_R}\right)}$ of the $T^{3\text{D}}$-spectrum of $\mathbf{Y}_{x_R}$ as
   $$\widehat{T^{3\text{D}}\left(\mathbf{Y}_{x_R}\right)} = \mathbf{W}_{x_R} T^{3\text{D}}\left(\mathbf{Z}_{x_R}\right),$$
   where the group $\mathbf{W}_{x_R}$ composed of the empirical Wiener shrinkage coefficients is defined by
   $$\mathbf{W}_{x_R} = \frac{\left(\check{T}^{3\text{D}}\left(\widehat{\mathbf{Y}}^{\text{ht}}_{x_R}\right)\right)^2}{\left(\check{T}^{3\text{D}}\left(\widehat{\mathbf{Y}}^{\text{ht}}_{x_R}\right)\right)^2 + \sigma^2}. \tag{7}$$
   Let us remark that here all multiplications and divisions of groups are element-by-element operations.

3. Inverse $T^{3\text{D}}$ transform (computed exactly as in Section 3.1.2).

3.2.3 Aggregation

As in Section 3.1.3, after processing all coordinates $x_R \in X$, we obtain local estimates on adaptive-shape neighborhoods from all the groups $\widehat{\mathbf{Y}}_{x_R}$, $x_R \in X$. In order to obtain a single *global* estimate $\hat{y}^{\text{wie}} : X \to \mathbb{R}$ defined on the whole image domain, we average together all the local estimates using a weighted averaging identical to (4). However, the weights w_{x_R} are now defined by

$$w_{x_R} = \frac{\sigma^{-2}}{\left\|\mathbf{W}_{x_R}\right\|_2^2 \left|\tilde{U}^+_{x_R}\right|}, \tag{8}$$

where the squared ℓ^2-norm of $\mathbf{W}_{x_R}$ is naturally computed as the sum of the square of each individual Wiener shrinkage coefficient contained in the group.

Image	σ	BM3D	P.SA-DCT	SA-BM3D
Barbara	10	**34.98** **0.9421**	33.50 0.9342	34.75 0.9419
	20	**31.78** 0.9054	30.00 0.8862	31.65 **0.9056**
	30	**29.81** 0.8687	28.10 0.8365	**29.81** **0.8709**
Lena	10	35.93 0.9166	35.58 0.9140	**35.95** **0.9175**
	20	33.05 0.8772	32.63 0.8718	**33.08** **0.8785**
	30	31.26 0.8449	30.86 0.8393	**31.33** **0.8474**
Montage	10	37.35 0.9679	37.12 0.9657	**37.57** **0.9682**
	20	33.61 0.9404	33.36 0.9362	**33.92** **0.9407**
	30	31.37 0.9114	31.06 0.9075	**31.70** **0.9138**
Cameraman	10	34.18 0.9319	33.98 0.9316	**34.35** **0.9331**
	20	30.48 0.8755	30.18 0.8752	**30.61** **0.8772**
	30	28.64 0.8375	28.24 0.8318	**28.72** **0.8384**

Table 1: PSNR (upper entries) and SSIM (lower entries) comparison between the proposed SA-BM3D, the BM3D, and the P.SA-DCT filters.

4. RESULTS

We present experimental results obtained with a preliminary version of the proposed SA-BM3D algorithm. The algorithm parameters are fixed for all experiments. Here we list some of the most

Image σ	10	20	30	40	50
Lena	35.95 0.9175	33.08 0.8785	31.33 0.8474	29.98 0.8198	28.50 0.7822
Cameraman	34.35 0.9331	30.61 0.8772	28.72 0.8384	27.29 0.8089	25.99 0.7664
Barbara	34.75 0.9419	31.65 0.9056	29.81 0.8709	27.69 0.8202	25.29 0.7218
Peppers	34.72 0.9283	31.22 0.8856	29.16 0.8501	27.67 0.8214	26.54 0.7961
House	36.85 0.9275	33.86 0.8781	32.12 30.74	30.74 0.8283	28.69 0.7897
Montage	37.57 0.9682	33.92 0.9407	31.70 0.9138	29.94 0.8884	28.24 0.8509
Boats	33.88 0.8893	30.81 0.8247	29.04 0.7773	27.70 0.7364	26.38 0.6856

Table 2: PSNR (upper entries) and SSIM (lower entries) results of the proposed SA-BM3D for various σ and test images.

important ones. The 1-D transform T^{1D} (part of the separable T^{3D}) was the 1-D Haar wavelet full-dyadic decomposition. The scales $H = \{1,2,3,4,5,6\}$ and $H = \{1,2,3,4\}$ are respectively used in the first and the second step. The maximum number of matched blocks $N_2 = 16$ in the first step and $N_2 = 32$ in the second. If there are fewer matches, only the best matching 2^j are kept, $j \in \mathbb{N}$, so to be able to apply the Haar transform.

In Table 1 we provide a comparison of the PSNR and the mean structural similarity index map [16] (SSIM) results of the SA-BM3D, BM3D and P.SA-DCT methods. A visual comparison is given in Figures 5 and 6, where we show enlarged details for *Barbara*, *Montage*, and *Cameraman*. In these figures we use a relatively high standard-deviation of the noise, $\sigma = 35$, in order to emphasize the differences in the results by each method. One can make the following observations.

- Textures in *Barbara* are preserved equally well by both the BM3D and the SA-BM3D — and significantly better as compared with the P.SA-DCT filter.
- Sharp edges in *Cameraman* are reconstructed equally well by both the P.SA-DCT and the SA-BM3D filters, while the BM3D filter introduces some ringing and blurring.
- The PSNR and SSIM results of the SA-BM3D are comparable or better than the ones of the BM3D and the P.SA-DCT filters. In particular, the problems of the P.SA-DCT filter in reconstructing textures, resulting in poor PSNR for *Barbara*, has been overcome by the SA-BM3D, whose results are comparable with the ones of the BM3D filter. An interesting result can be seen in Table 1 for *Barbara* in the case of $\sigma = 20, 30$ and in Figure 5 for $\sigma = 35$; in these cases the BM3D performs best in PSNR but the SA-BM3D is best in SSIM[4].

In addition, in Table 2 we give PSNR and SSIM results of the SA-BM3D for a broader set of test images and various values of σ.

5. CONCLUSIONS

As a simultaneous generalization of the BM3D and the Pointwise SA-DCT filters, the proposed denoising method exploits both:

- nonlocal image modeling realized by *finding and grouping* of similar image neighborhoods (as in the BM3D filter);
- pointwise adaptive anisotropic estimation realized by grouping of *adaptive-shape image neighborhoods* (as in the P.SA-DCT filter).

The developed algorithm was shown in Section 4 to inherit the strengths of both methods (i.e. BM3D's good reconstruction of textures and regular image structures and P.SA-DCT's good reconstruction of sharp edges and image singularities) and at the same compensating for their deficiencies (i.e. P.SA-DCT's ineffectiveness for textures and the BM3D's blurring and detail loss around sharp curved edges and singularities).

[4] A discussion about the subjective perceptual quality of the P.SA-DCT, with particular emphasis on the restoration of textures, can be found in [4].

REFERENCES

[1] A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Modeling and Simulation*, vol. 4, no. 2, pp. 490–530, 2005.

[2] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, August 2007.

[3] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. on Image Process.*, vol. 15, no. 12, pp. 3736–3745, December 2006.

[4] A. Foi, *Pointwise Shape-Adaptive DCT Image Filtering and Signal-Dependent Noise Estimation*. D.Sc.Tech. Thesis, Institute of Signal Processing, Tampere University of Technology, Publication 710, 2007.

[5] A. Foi, V. Katkovnik, and K. Egiazarian, "Pointwise Shape-Adaptive DCT for high-quality denoising and deblocking of grayscale and color images," *IEEE Trans. Image Process.*, vol. 16, no. 5, pp. 1395–1411, May 2007.

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

[7] J. Guerrero-Colon and J. Portilla, "Two-level adaptive denoising using Gaussian scale mixtures in overcomplete oriented pyramids," in *Proc. IEEE Int. Conf. Image Process.*, vol. 1, Genova, Italy, September 2005.

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

[9] V. Katkovnik, A. Foi, K. Dabov, and K. Egiazarian, "Spatially adaptive support as a leading model-selection tool for image filtering," in *Proc. 2008 Workshop Inf. Theoretic Methods in Science and Engineering, WITMSE 2008*, Tampere, Finland, August 2008.

[10] V. Katkovnik, A. Foi, K. Egiazarian, and J. Astola, "Directional varying scale approximations for anisotropic signal processing," in *Proc. European Signal Process. Conf.*, Vienna, Austria, September 2004, pp. 101–104.

[11] V. Katkovnik, K. Egiazarian, and J. Astola, *Local Approximation Techniques in Signal and Image Processing*. SPIE Press, 2006, vol. PM157.

[12] P. Kauff and K. Schüür, "Shape-adaptive DCT with block-based DC separation and ΔDC correction," *IEEE Trans. Circuits Syst. Video Tech.*, vol. 8, no. 3, pp. 237–242, March 2002.

[13] C. Kervrann and J. Boulanger, "Optimal spatial adaptation for patch-based image denoising," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 2866–2878, October 2006.

[14] S. Lyu and E. Simoncelli, "Modeling multiscale subbands of photographic images with fields of Gaussian scale mixtures," *to appear in IEEE Trans. Patt. Analysis and Machine Intell.*, 2008.

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

[16] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, April 2004.

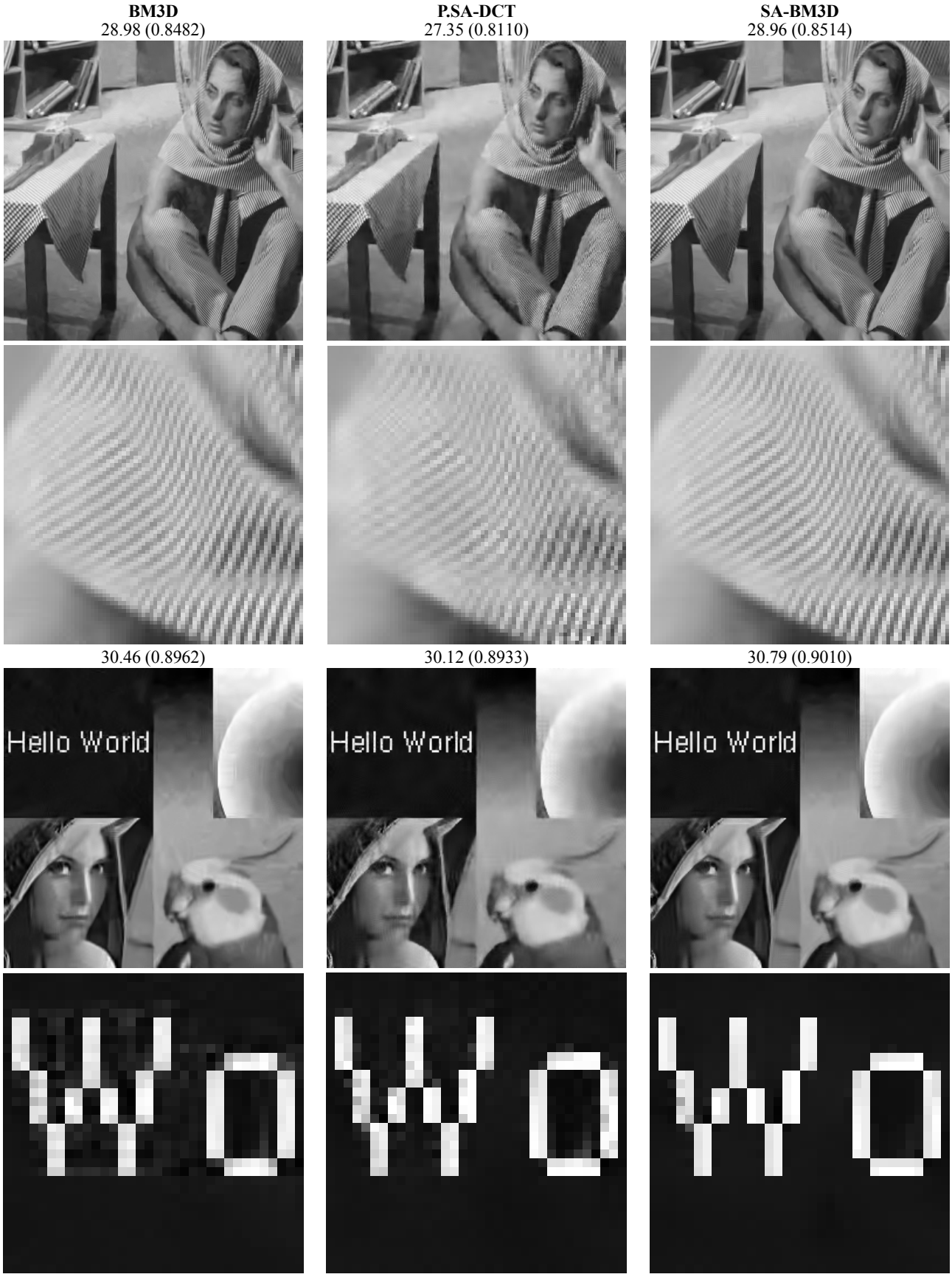

Figure 5: Comparison between denoised *Barbara* and *Montage* by the BM3D (left), the SA-DCT (center), and the SA-BM3D (right) filters. The numbers above the images are the corresponding PSNR and SSIM (in parentheses) values (calculated on the whole image). The standard deviation of the noise in the input images is $\sigma = 35$.

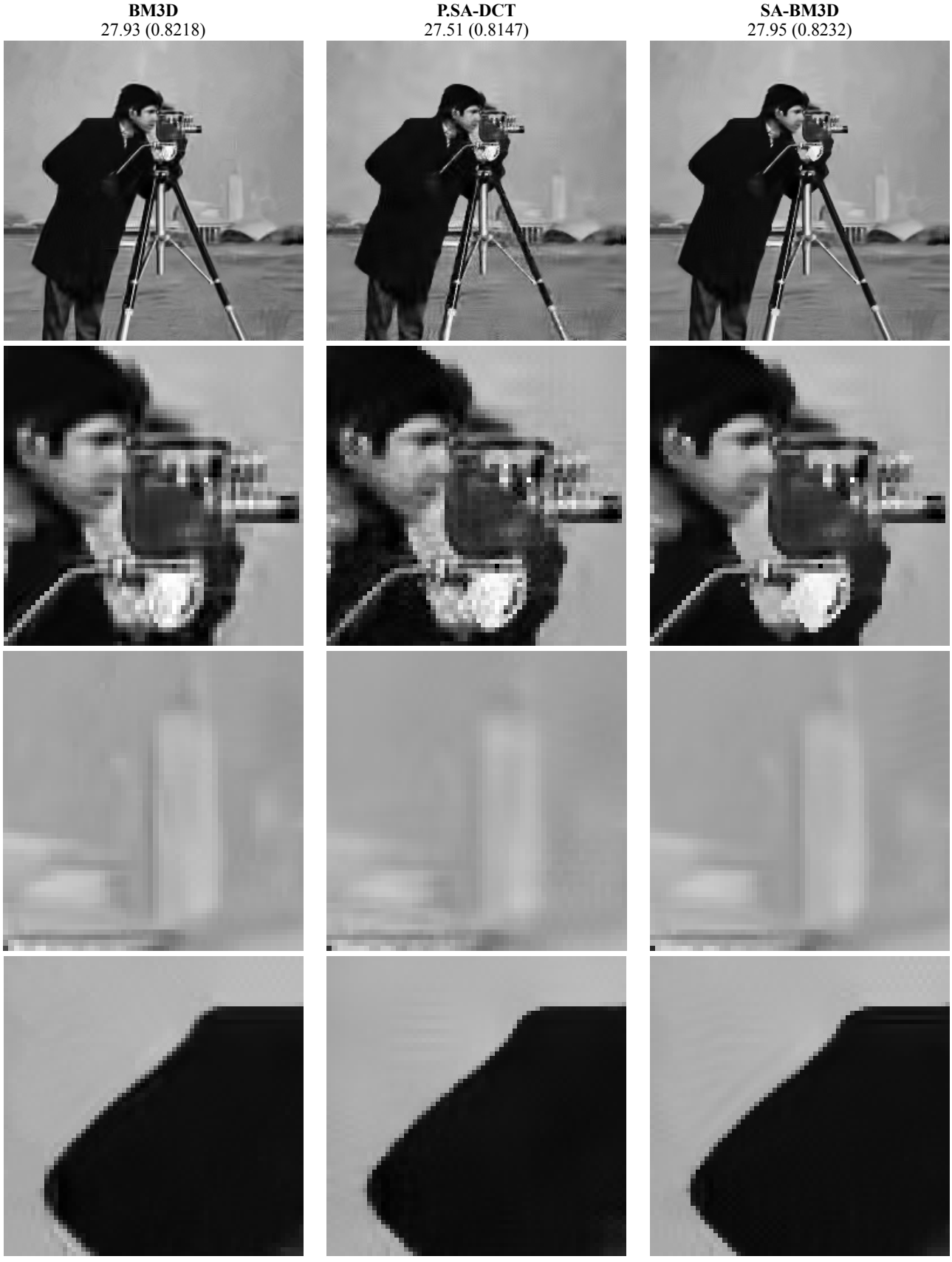

Figure 6: Comparison between denoised *Cameraman* by the BM3D (left), the SA-DCT (senter), and the SA-BM3D (right) filters. The numbers above the images are the corresponding PSNR and SSIM (in parentheses) values (calculated on the whole image). The standard deviation of the noise in the input image is $\sigma = 35$.