

On the comparison of different kernel functionals and neighborhood geometry for nonlocal means filtering

José I. de la Rosa^{1,2} · Jesús Villa-Hernández¹ ·
Joaquín Cortez² · Hamurabi Gamboa¹ · José G. Arceo¹ ·
Efrén González¹

Received: 24 May 2016 / Revised: 22 December 2016 / Accepted: 27 December 2016 /
Published online: 10 January 2017
© Springer Science+Business Media New York 2017

Abstract The present work proposes a review and comparison of different Kernel functionals and neighborhood geometry for Nonlocal Means (NLM) in the task of digital image filtering. Some different alternatives to change the classical exponential kernel function used in NLM methods are explored. Moreover, some approaches that change the geometry of the neighborhood and use dimensionality reduction of the neighborhood or patches onto principal component analysis (PCA) are also analyzed, and their performance is compared with respect to the classic NLM method. Mainly, six approaches were compared using quantitative and qualitative evaluations, to do this an homogeneous framework has been established using the same simulation platform, the same computer, and same conditions for the initializing parameters. According to the obtained comparison, one can say that the NLM filtering could be improved when changing the kernel, particularly for the case of the Tukey kernel. On the other hand, the excellent performance given by recent hybrid approaches such as NLM SAP, NLM PCA (PH), and the BM3D SAPCA lead to establish that significantly improvements to the classic NLM could be obtained. Particularly, the BM3D SAPCA approach gives the best denoising results, however, the computation times were the longest.

Keywords Image filtering · Nonlocal means · Kernel functionals · Sparse learning techniques · Collaborative filtering

✉ José I. de la Rosa
ismaelrv@ieee.org

¹ Unidad Académica de Ingeniería Eléctrica, Universidad Autónoma de Zacatecas, Antiguo Camino a la Bufa No. 1, Col. Centro., Zacatecas, C. P. 98000, México

² Depto. de Ingeniería Eléctrica y Electrónica, Instituto Tecnológico de Sonora, Campus Náinari, Av. Antonio Caso No. 2266, Col. Villa ITSON, Cd. Obregón, Sonora, C.P. 85130, Mexico

1 Introduction

The image denoising methods are at the basis of digital image processing and still, there are challenging problems to construct robust filters in applications where the noise has very large standard deviation or where it is of multiplicative nature such as the speckle type noise [15, 37]. Since 2005, the Nonlocal Means (NLM) filtering has gained some popularity and credibility, because this type of filters deals with the preservation of structure and objects into a digital image [2]. The “method noise” is the core of the mathematical analysis of such nonlocal means filters where an error is defined as the difference between a noisy image and its denoised version.

The image filtering or denoising, is a particular task of restoration or recuperation approaches of an image to its original condition given a degraded image. The restoration passes by reverting the effects caused by a distortion function which must be estimated in most of the practical cases. In fact, the degradation characteristic is a crucial information and it must be supposed known or estimated during the inversion procedure. Typically, this is a point spread function which can be linked with the probability distribution of the noise contamination ($n \sim p(n)$). Thus, a global image formulation model could be given by:

$$y = F(x) + n, \quad (1)$$

where, $F(x)$ is a functional that could take for instance, two forms: $F(x) = x$ and $F(x) = Hx$, being H a linear operator which models the image degradation. All variables presented along the text are, x : which represents an image to be estimated, y : represents the observed image with additive white noise n and/or distorted by H , and \hat{x} : is the estimator of x with respect to data y .

The classic exponential kernel function is used in most of the cases as a weighting function in NLM methods [2–5] to deal with digital image filtering as it can be seen in (8). Moreover, the performance of the NLM depends on an optimal choice of the bandwidth parameter h , and other related parameters such as the size of the neighborhood, and the search region (geometry of patches). An optimal selection of the bandwidth parameter h is a difficult task, this problem has been interestingly solved in the works of Van De Ville and Kocher [43, 44], and Talebi [39], by using the Stein’s unbiased risk estimate (SURE) and locally adaptive procedures, also one can propose an empirical fixed approximation based on experimental results. Moreover, since h depends on the noise level to have a good performance of the NLM methods, in practice, it is important to estimate the noise distribution or some of its statistical properties, recently some works have reported some performing approaches in this important task [6, 9, 34, 35].

On the other hand, there are some other kernel functions which have other equivalent parameters [19] and they may be used in the NLM context as robust weighing functions as illustrated in [20, 22, 42]. More over, the performance of NLM could be improved by using recent hybrid algorithms which are based on the change of the geometry of the neighborhood of noisy pixels and performing a collaborative filtering, here the notion of block-wise or patch-wise filtering plays an important role in NLM, since the methodology can be improved as shown in works of Deledalle et al. [14, 15], and the successful BM3D method proposed by Dabov et al. [7, 8]. In the present paper we propose the review and performance comparison of different Kernel functionals and hybrid methods. First, one is interested to answer the following questions: what happens with the performance of NLM methods if one changes the kernel structure? And a second question is, what happens when one changes

the geometry of the searching region for NLM? To answer both questions, some comparisons were performed with respect to the classic NLM method, to do this an homogeneous framework has been established using the same simulation platform, the same computer, and same conditions for the initializing parameters, such as the seed to generate the noise random samples which will be added to free-noise database images.

The paper is organized according to the following sections: Section 2 describes the general presentation of the Nonlocal Means filtering technique. The nonparametric estimation and its connection to NLM is shown in Section 3 where the notion of kernel and robust weighting functionals is presented and repositioned for the task of NLM filtering. In Section 4, a summarized presentation of patch-wise based NLM is boarded. In Section 5, the results of NLM filtering are discussed for several approaches and some illustrative results are shown to answer the questions on the use of robust kernels and changing geometry of the searching region into the NLM framework. Finally, in Section 7 some concluding remarks are given.

2 Nonlocal means (NLM) filter

Let us consider the following observation model, which is a particular case of (1)

$$y(i) = x(i) + n(i), \quad (2)$$

where $y(i)$ is the observed value, $x(i)$ is the true value, and $n(i)$ is the noise perturbation at a pixel i . The classic way to model the effect of noise on a digital image is the additive white noise, for example, Gaussian noise where $n(i) \sim \mathcal{N}(0, \sigma_n^2)$ (AWGN).

Buades et al. [2–5] have defined the denoising method D_h according to the following equation:

$$y = D_h(y) + n(D_h, y), \quad (3)$$

where y is the noisy image and h is a filtering parameter, which usually depends on the standard deviation of the noise. Ideally, $D_h(y)$ is smoother than y and $n(D_h, y)$ models the realizations of a white noise.

Definition 1 *Let y be any image and D_h a denoising operator depending on h . Then, the method noise of y is defined by the following image difference*

$$n(D_h, y) = y - D_h(y). \quad (4)$$

Moreover, in [4] the authors establish that a good denoising method produces a method noise $n(D_h, y)$ that is distributed as close as possible to the original noise distribution according to following three principles:

Principle 1: For every denoising algorithm, the method noise must be zero if the image contains no noise and should be in general an image of independent zero-mean random variables.

Principle 2: Noise-to-noise principle. A denoising algorithm must transform a white noise image into a white noise image (with lower variance).

Principle 3: Statistical optimality. A generalized neighborhood filter is optimal if it finds for each pixel i all and only the pixels j having the same model as pixel i .

The previous definition, is at the core of the proposed NLM method by Buades et al. [2–5], which follows the next integration formula

$$\hat{x}(i) = NL(y)(i) = \frac{1}{C(i)} \times \int_{\Omega} \exp\left(-\frac{(G_a * |y(i + \cdot) - y(j + \cdot)|^2)(0)}{h^2}\right) y(j) dj \tag{5}$$

where $i \in \Omega \subset \mathbb{R}^2$, G_a is a Gaussian kernel with a standard deviation a , h is a bandwidth which acts as a filtering parameter, and thus the choice of this parameter is important, and in practice it is estimated in function to the noise variance σ_n^2 or standard deviation $h = k_0\sigma_n$, with a constant k_0 [6, 9, 34, 35] or estimated according to [43, 44]. The value of $C(i)$ is a normalizing constant such that

$$C(i) = \int_{\Omega} \exp\left(-\frac{(G_a * |y(i + \cdot) - y(j + \cdot)|^2)(0)}{h^2}\right) dj,$$

and

$$(G_a * |y(i + \cdot) - y(j + \cdot)|^2)(0) = \int_{\mathbb{R}^2} G_a(t) |y(i + t) - y(j + t)|^2 dt.$$

One could say that $NL(y)(i)$ is the denoised value at the i -th position, and it is the mean value of all pixels whose Gaussian neighborhood looks like the neighborhood of i -th pixel.

2.1 The classic discrete proposition

According to the previous introduction of the NLM method (5), this is approximated by the following sum for the discrete images case (discrete grid $i \in \mathcal{I} \subset \mathbb{Z}^2$)

$$\hat{x}(i) = \frac{1}{C(i)} \sum_{j \in \mathcal{I}} w(i, j) y(j), \tag{6}$$

where the wights $\{w(i, j)\}_j$ depend on the similarity between the pixels i and j and satisfy the conditions: $0 \leq w(i, j) \leq 1$, $\sum_j w(i, j) = 1$, and $C(i)$ is a normalizing factor given by

$$C(i) = \sum_{j \in \mathcal{I}} w(i, j), \tag{7}$$

being \mathcal{I} the searching region around i and $w(i, j)$ is a weighting function, that compares the neighborhoods around pixels i and j , such that

$$w(i, j) = \exp\left(-\frac{\|y(\mathcal{N}_i) - y(\mathcal{N}_j)\|_{2,a}^2}{h^2}\right), \tag{8}$$

where \mathcal{N}_i defines a neighborhood system on \mathcal{I} .

Definition 2 A neighborhood system on \mathcal{I} is a family $\mathcal{N} = \{\mathcal{N}_i\}_{i \in \mathcal{I}}$ of subsets of \mathcal{I} such that for all $i \in \mathcal{I}$,

- 1) $i \in \mathcal{N}_i$,

$$2) \quad j \in \mathcal{N}_i \Rightarrow i \in \mathcal{N}_j.$$

The subset \mathcal{N}_i is called the neighborhood, similarity window of i , or patch window.

The neighborhoods or similarity windows have different sizes and shapes, the most common shape is a squared window of fixed size. The restriction of y to a neighborhood \mathcal{N}_i is denoted by

$$y(\mathcal{N}_i) = (y(j), j \in \mathcal{N}_i). \quad (9)$$

2.2 Some new trends

Since the introduction of the NLM approach proposed by Buades et al. [2, 3, 5] for image filtering, several generalizations or alternatives have been proposed in the literature, hybrid approaches such as combination of NLM and Wavelets [28, 45, 46], and also more general NLM and Linear Transforms, NL Variational Methods [37], NL anisotropic patches [31], the focusing on the calculations of an appropriate or optimal bandwidth selection [39, 43, 44], the appropriate size or geometry of the patches or neighborhoods of analysis [14], and the change of the kernel or weighting function to be used [13, 14, 20, 22, 38, 42], since the most of time the exponential function is used. Also, the notion of self-similarity or redundancy has been explored to construct methods to accelerate the NLM approach such as the pre-selection of the contributing neighborhoods based on average value and gradient, average and variance or higher-order statistical moments [38], Milanfar [33], cluster tree arrangement, and singular value decompositions [15], such as Principal Component Analysis. Also the computation of the distance measure between different neighborhoods can be optimized using the fast Fourier transform [14, 29] or a moving average filter obtaining fast algorithms. One of the best alternatives nowadays is of course, the hybrid method of sparse 3-D transformation with collaborative filtering (Block Matching 3-D–BM3D) [7, 8]. Other recent propositions concern Bayesian patch-based methods [27], or the adaptive penalized NLM [37] which search to have a similar performance or better than the BM3D method. More over, these methods have been adapted to obtain robust image filtering where non-Gaussian or multiplicative noise is present.

3 Connections of nonparametric estimation and NLM

According to the work of Takeda et al. [38], and Milanfar [33], and from (1) it is possible to obtain a kernel regression formulation such that an image estimator is given by the following expression,

$$\hat{F}(x_i) = \frac{\sum_{j \in \mathcal{I}} K_h(y(\mathcal{N}_i) - y(\mathcal{N}_j))y(j)}{\sum_{j \in \mathcal{I}} K_h(y(\mathcal{N}_i) - y(\mathcal{N}_j))}, \quad (10)$$

also according to the consistency theorem given by Buades [3], the previous equation corresponds also to the NLM method, where classically the Nadaraya–Watson kernel assumes an exponential structure, such as in (8), where one can establish that

$$w(i, j) = K_h(y(\mathcal{N}_i) - y(\mathcal{N}_j)). \quad (11)$$

This last equation let us to connect the nonparametric estimation framework with the NLM method (see also [26]), where a kernel structure is necessary to obtain an estimate version $\widehat{p}_{m,h}(z)$ (empirical distribution) of a distribution $p(z)$ where z is a vector of independent and identically distributed random variables of size m . For the 1-dimensional case, let the following expression denote such estimators:

$$\widehat{p}_{m,h}(z) = \widehat{p}_{m,h}(z|z_1, \dots, z_m) = \frac{1}{m} \sum_{i=1}^m K_h(z - z_i). \quad (12)$$

This expression assumes the hypothesis that $p(z)$ is symmetric, two times differentiable and positive, indeed, it is also assumed that $K_h(\cdot)$ is a kernel weighted function which satisfies some imposed conditions treated in the works of Berline [1], Devroye [16–18], Loader [30] and Masry [32]. The bandwidth $h = h_m$ is given in function of both the sample size m and the standard deviation of z , this parameter could be considered as a sequence of positive numbers that must satisfy: $h_m \rightarrow 0$ and $mh_m \rightarrow \infty$ when $m \rightarrow \infty$. The strong uniform consistency of $\widehat{p}_{m,h}(z)$ and its convergence toward $p(z)$, depend on a convenient procedure of bandwidth selection. Berline and Devroye have made a complete study, where they have compared several classic and plug-in techniques [1], [17]. A simple and faster procedure which has been retained for this work is the technique proposed and developed by Terrell. It has been shown that under reasonable conditions of symmetry and for monomodal distributions this procedure is consistent, the complete conditions to assure the global consistency, efficiency and convergence are given in [40, 41], which is very similar to the estimate technique selected by You et al. in [45, 46] in the framework of hybrid NLM and Wavelets. The nonparametric framework for denoising signals [12, 26, 33] seems to be a parallel tool with respect to NLM techniques, where for instance the exponential kernel structure is the most used.

A function of the form $K_h(z)$ is assumed as a fixed kernel $K_h(z) = 1/(h^d)K(z/h)$, where $h > 0$, this parameter is called the kernel bandwidth (smoothing factor). The fundamental problem in kernel density estimation lies in both the selection of an appropriate value for h and the selection of the kernel structure. The choice of $K(z)$ could depend on the smoothness (regularity) of $p(\cdot)$. Two different nonparametric schemes are revisited in this section to connect nonparametric to NLM methods. The first one uses the exponential kernel, which has proved to give good performance when h is selected by using the over-smoothed principle introduced by Terrell. And the second uses a kernel obtained from the class of Hilbert kernels proposed in [19]. It avoids the bandwidth h selection and its performance depends on other parameters, which selection is easier (parameters d and k are defined in Section 3.2).

3.1 Exponential kernel with h optimally estimated

Among the different classic kernels [1], the Gaussian kernel is the most utilized in nonparametric estimation due to its regularity and symmetric properties (Gaussian decay) and it leads to an easy to implement estimator. The following expression resumes this estimator by a sum of exponential functions:

$$\widehat{p}_{m,h}(z) = \frac{1}{mh\sqrt{2\pi}} \sum_{i=1}^m \exp\left(-\frac{(z - z_i)^2}{2h^2}\right). \quad (13)$$

In such a case, and considering that a fixed kernel structure has been chosen, Terrell [40] proposes to use an over-smoothed bandwidth h that corresponds to:

$$h_0 = 3 \left(\frac{1}{2\sqrt{\pi}(35)} \right)^{\frac{1}{5}} \sigma m^{-\frac{1}{5}},$$

this bandwidth value guarantees the minimization of the Mean Integrated Squared Error (MISE), σ is the standard deviation of the sample z , and $\int K(z)^2 dz = \frac{1}{(2\sqrt{\pi})}$. Under mild conditions, the kernel density estimates based on the over-smoothing principle are consistent and for sufficiently large sample sizes m , they will display all information present in the underlying variables density $p(z)$. This way to approximate h will be used into the NLM classic method, where the h_0 depends adaptively on the search regions of the image, for relatively small homogeneous areas [35] ($\mathcal{L} = [-5, 5] \times [-5, 5], [-7, 7] \times [-7, 7]$) obtaining an acceptable noise variance estimation and standard deviation for each search region.

3.2 Hilbert kernel

Another class of kernel density estimates is called the Hilbert kernel estimate. In this case, the $K_h(z) = 1/(h^d)K(z/h)$ is considered equivalent to $K(u) = 1/\|u\|^d$, where the smoothing factor h is canceled obtaining:

$$\hat{p}_m(z) = \frac{1}{m} \sum_{i=1}^m \frac{1}{\|z - z_i\|^d}. \tag{14}$$

The Hilbert estimates are viewed as an universally consistent density estimate whose expected performance (L_1, L_∞ , pointwise) is monotone in m (at least in theory) for all densities. The consistency of this class of estimators is proved in [18](see theorem 2). The Hilbert density estimate of order k ($k \geq 1$) is a redefined subclass that avoids the infinite peaks produced during the estimation, the value considered most of the time is $k = 2$, giving the following expression for $\hat{p}_m(z)$:

$$\hat{p}_m(z) = \sqrt{\frac{4}{V_d^2 \pi m(m-1) \log m} \sum_{1 \leq i < l \leq m} \frac{1}{\text{Den}_{i,l}}}, \tag{15}$$

where $\text{Den}_{i,l} = \|z - z_i\|^{2d} + \|z - z_l\|^{2d}$ and V_d is the volume of the unit ball in \mathbb{R}^b . This last expression is also called Cauchy density estimate, due to its similarity to the multivariate Cauchy density, $\|\cdot\|$ denotes the L_2 metric on \mathbb{R}^d . Finally, it is assumed that $\hat{p}_m(z) \rightarrow p(z)$ at least in probability for almost all z .

From the previous nonparametric estimator, and to answer the first question, about what happens if one changes the exponential kernel by another (replacing (8) by (16)), it is now proposed to use as a weighting function for the NLM method the Hilbert kernel, that is:

Proposition 1 Hilbert kernel:

$$w(i, j) = \frac{1}{\|y(\mathcal{N}_i) - y(\mathcal{N}_j)\|_{2,a}^d}, \tag{16}$$

where one have experienced in other works with integer values for $d = 1, 2, 3, 4$ [10–12].

This proposition, is given in the same sense of the robust filtering proposition made by Dinesh et al [20], where some interesting results are discussed. More over, one notes that this is a generalization of the BLUE function discussed in [22], for $d = 2$ where $w(i, j) = 1/h^2$ when $\|y(\mathcal{N}_i) - y(\mathcal{N}_j)\|_{2,a}^2 \leq h$.

3.3 Other robust kernels

Also the works of Goossens et al. [22] and Tian et al. [42] give some interesting kernel propositions, in the case of [42] with poor performance in the filtering, which was due to an erroneous consideration on $w(i, j)$ and h which has been corrected in this work to perform the comparison given in the Section 5. The following two kernels were also replaced as weighting functions for the NLM method (changing the (8) by (17) and (18) respectively), where the value of λ reported in [42] was substituted by h^2 . Specifically, these two functions were taken into account because they have good performance characteristics (see also the modified bi-square proposed in [22]),

Proposition 2 The Tukey or bi-square function:

$$w(i, j) = \left(1 - \left(\frac{\|y(\mathcal{N}_i) - y(\mathcal{N}_j)\|_{2,a}^2}{h^2} \right) \right)^2, \quad (17)$$

for $0 < \|y(\mathcal{N}_i) - y(\mathcal{N}_j)\|_{2,a}^2 \leq h$.

Proposition 3 The Andrews or Wave function:

$$w(i, j) = \frac{\sin\left(\frac{\pi \|y(\mathcal{N}_i) - y(\mathcal{N}_j)\|_{2,a}}{h}\right)}{\pi \|y(\mathcal{N}_i) - y(\mathcal{N}_j)\|_{2,a}/h}, \quad (18)$$

for $0 < \|y(\mathcal{N}_i) - y(\mathcal{N}_j)\|_{2,a}^2 \leq h$.

4 Neighborhood geometry and dimensionality reduction in NLM

Some of the recent hybrid algorithms are based on the change of the geometry of the neighborhood and the collaborative filtering, this is lead by using dimensionality reduction of the neighborhood or patches onto principal component analysis (PCA) improving the performance of the classic NLM, since the methodology can model far better the geometries and textures as shown in works of Deledalle et al. [14] (NLM with Shape Adaptive Patches–SAP), [15] (NLM PCA), were promising results are given and are comparable with those of the successful BM3D method proposed by Dabov et al. [7], which also has adopted the philosophy of adapting shapes into a PCA [8] improving its own results (BM3D SAPCA). In [22] the authors also propose PCA decomposition, and uses a post-processing filter that seems to give competitive results with respect to BM3D method (in the present work we only change the proposed kernels in Section 3.3). The core in this new type of NLM based methods is the use of orthogonal over complete dictionaries combined with sparse learning techniques where the dictionaries are learned directly from the noisy image by using a PCA decomposition of patches, this stage is also complex since one is faced to the selection of the best dictionaries. All these methods also require the knowledge of σ_n to calculate the

optimal bandwidth h . The filtering task is thus lead by denoising an image block-wise or patch wise instead of a single pixel, that means the following

$$\hat{x}_b(i) = \frac{\sum_{j \in \mathcal{I}} \sum_{k=-K}^K b(k)w(i+k, j+k)y(j)}{\sum_{j \in \mathcal{I}} \sum_{k=-K}^K b(k)w(i+k, j+k)}, \quad (19)$$

where $\hat{x}_b(i)$ are multiple estimations of the i -th pixel based on a block or patch based NLM (with K overlapping blocks), and $b(k)$ is an additional weighting function to aggregate the different estimates.

On the other hand, the usual geometry for the patches is a square, if one changes the shape of the patch taking advantage of the local geometry of the image, one arrives to construct anisotropic patches with the benefit of directionality and best geometry representation as discussed in [31]. This is the main idea in [8, 14], where some type of shapes have been proposed (Disks, Pie slices, and Bands). The final task into these approaches is to carry out a procedure called aggregation, where several estimators are combined giving the best solution (computing several times the NLM, this is made for each shape). The performance of these methods is great, since the aim is to model adequately the texture content and at the same time model the edges with high contrast. For example, in the case of BM3D SAPCA [8], the processing is carried out according to the following stages: Shape adaptive grouping, obtaining the shape by 8 directional filters with effective sparsity of the image data and finding similar blocks; then PCA basis are obtained and used into the collaborative filtering obtaining a 3-D transformation, followed by a Shrinkage task and inverting the transformation. An aggregation task is performed to obtain the final denoised image.

5 Some simulation results and comparison

Some results were obtained conducting two different experiments. For both experiments, Gaussian noise random samples were added to free-noise database images (the (2) has been implemented) obtaining noisy images which were denoised using the following approaches:

Approach 1 Corresponds to the classic NLM method (software elaborated by Manjon and Buades downloaded from [25]), but taking into account that the bandwidth h values chosen to parameterize the NLM algorithm depends directly on the true σ_n of the noise (here some values were chosen according to those reported in [43], but for the fixed values $h = 0.7\sigma_n$, $h = \sigma_n$, and $h = 1.5\sigma_n$).

Approach 2 Since in practice, for real acquired images the true value of σ_n is often unknown, the second approach considers to still use the classic NLM algorithm but estimating optimally the value for $h = k_0 h_0$ (where the variance and standard deviation are also estimated using maximum likelihood estimators, for each search region \mathcal{I} (see Section 3.1)).

Approach 3 For a third approach, (16), (17), and (18) were replaced by (8), and the role of parameter h was changed by d for the Hilbert kernel, where the noisy images were filtered by using the following integer values $d = 2, 3, 4$ for the case of the Hilbert kernel, while for the case of the Tukey and Andrews kernels the value $h = \sqrt{6}\sigma_n$ which is nearest to the value proposed in [22].

Approach 4 In this case, the approach used is NLM with Shape Adaptive Patches (NLM SAP) which was downloaded from the web page of Deledalle [24], and corresponds to an hybrid patch based algorithm. This method uses a trapezoidal kernel and only three pie slices as shapes, obtaining a fast version proposed by authors (in this case $h^2 = 2\sqrt{8\sigma_n^4|S|}$, where S is an equivalent size of the shape).

Approach 5 Other hybrid approach is the patch based PCA: local vs. global (NLM PCA), proposed in [15], here authors compare patch global PCA (PG), hierarchical PCA (PH), and local PCA (PL), the software was also downloaded from the web page of Deledalle [24] and (PH) results were considered for our comparison purposes. Almost all the default parameters to simulate this approach have been preserved, and only the values for M_{min} were replicated according to the value of σ_n (in the function `PHPCA_best_params.m`).

Approach 6 The final simulated approach corresponds to the method BM3D shape-adaptive (SA) PCA proposed in [8], the software was downloaded from the web page of Foi [23]. The default parameters to simulate this approach have been preserved.

From the comparison of approaches 1, 2, and 3, the results let us to answer to the first question asked in the introduction section, and results obtained with approaches 4, 5, and 6 let us to answer the second question. In the first experiment, the obtained results are compared with respect to some other reported results in literature. From a classic database testing images only Barbara, Lena, Cameraman, Mandril and Boats were used, trying to corroborate results reported in other references cited in the present paper. The Gaussian noise has zero mean, with different variance values, with standard deviations $\sigma_n = 10, 15, 20, 25, 30, 40$ (medium to high level of noise). In the second experiment, also some results were obtained for the complete database TID2008 (25 images) which has been used in [35], in this experiment the values for the standard deviation of noise are $\sigma_n = 0, 1, 3, 5$ (noiseless to low level of noise). For both experiments an objective and a subjective comparisons have been performed. In Sections 5.1 and 5.2 the Peak Signal to Noise Ratio (PSNR) has been considered as a measure to quantify the performance of all compared approaches (one can also use other quantification like the Structural Similarity (SSIM)), while in Section 5.3 some comments are made concerning a subjective evaluation qualifying the visual perception of some denoised images. Also, all results were obtained using MATLAB version 2010a, a personal computer with AMD A10 APU Processor, with 8 GB of RAM and bus of 64 bits, and the random seed code in MATLAB was `randn('seed', 2)`, and the same function for the PSNR.

5.1 PSNR results for some classic test images

Table 1 shows the PSNR values obtained when filtering or denoising the Barbara test image. Here, it is compared in an objective way the performance level of restoration versus the level of noise. Also, in Table 2 some similar performance results were obtained for Lena test image. In both tables, the six implemented approaches previously described are compared with respect to other two approaches reported in literature, particularly those results reported by Lin [28] (in Tables 1 and 2 appears as NLM Wiener-Wavelets) and by You [46] (in Tables 1 and 2 appears as NLM Wavelet). One can see from Tables 1 and 2, that obtained results using approaches 1, 2, and 3 are improved when changing the kernel, particularly for the case of Tukey kernel. The PSNR results obtained by the NLM Wiener-Wavelets and NLM Wavelet domain perform little bit better for filtering high level of noise, this is due in part to the goodness of these hybrid approaches which combine the image analysis in

Table 1 PSNR results in dB for evaluating some NLM methods (Barbara image)

Method \ σ_n	10	15	20	25	30	40
NLM Buades						
$h = 0.7\sigma_n$	32.97	30.95	29.37	28.21	27.37	26.00
$h = \sigma_n$	33.55	31.48	29.83	28.51	27.39	25.75
$h = 1.5\sigma_n$	32.56	30.14	28.33	26.93	25.86	24.41
NLM h est.						
$h = 1.5h_0$	29.82	29.03	28.18	27.34	26.54	25.31
$h = h_0$	32.12	30.40	28.87	27.56	26.42	24.65
NLM Hilb.						
$d = 2$	30.87	29.33	28.08	27.06	26.23	24.90
$d = 3$	32.57	30.69	29.16	27.89	26.82	25.25
$d = 4$	32.95	30.80	29.15	27.84	26.79	25.07
NLM other						
[22, 42] Tukey	33.82	32.11	29.82	28.53	27.43	25.67
[22, 42] Wave	33.76	31.86	29.42	27.97	26.74	24.90
NLM W-W						
[28]	–	31.17	30.47	29.55	28.33	–
NLM W						
[45, 46]	–	–	30.27	–	28.27	–
NLM SAP						
[14]	33.79	32.10	30.87	29.62	28.24	26.51
NLM PCA						
(PH) [15]	34.55	32.32	30.92	29.68	28.56	26.98
BM3D PCA						
[8]	34.96	33.19	31.89	30.87	30.01	28.50

the spatial and in the frequency domains (the centered lines in Tables 1 and 2 indicate that [28] and [45, 46] have not reported PSNR results for the corresponding σ_n). More over, the best performance is evidenced by the approaches 4, 5, and 6, changing the geometry of the neighborhood and using the notion of patch wise filtering (collaborative filtering using dimensionality reduction of the neighborhood or patches onto principal component analysis). Particularly, the method BM3D SAPCA gives the best performance, whereas the methods NLM SAP and NLM PCA (Patch based Hierarchical–PH) give competitive results and are valuable since their computation times were the fastest (see last column of the Table 6, for the case of the database TID2008).

In the same way, in Tables 3, 4, and 5 some other comparatives are shown with respect to other test images such as Cameraman, Madril and Boats. In these cases, the performance of the six approaches is similar comparing with Tables 1 and 2. Moreover, the approach NLM SAP gives the better performance for values of $\sigma_n > 20$ (high level of noise) for Cameraman and Boats images, whereas the performance of the BM3D SAPCA is generally the best. The Mandril image case is interesting since it is rich of texture content and the approaches performs a little bit different concerning the h value, for example NLM classic method performs better for $h = 0.7\sigma_n$, obtaining a PSNR values near to those of the Tukey Kernel and NLM SAP. Moreover, in the case of the classic NLM method (approach 1), it is

Table 2 PSNR results in dB for evaluating some NLM methods (Lena image)

Method \ σ_n	10	15	20	25	30	40
NLM Buades						
$h = 0.7\sigma_n$	34.08	32.08	30.64	29.32	28.25	26.60
$h = \sigma_n$	35.05	33.21	31.42	30.24	29.26	27.66
$h = 1.5\sigma_n$	34.22	32.30	30.92	29.85	29.00	27.72
NLM h est.						
$h = 1.5h_0$	32.31	31.31	30.28	29.31	28.40	26.94
$h = h_0$	33.38	31.14	29.45	28.16	27.10	25.44
NLM Hilb.						
$d = 2$	32.85	31.47	30.35	29.40	28.58	27.19
$d = 3$	33.94	32.12	30.63	29.38	28.41	27.03
$d = 4$	34.01	31.81	30.11	28.95	28.01	26.54
NLM other						
[22, 42] Tukey	35.28	33.00	31.51	30.38	29.43	27.93
[22, 42] Wave	35.13	32.78	31.25	30.10	29.15	27.71
NLM W-W						
[28]	–	32.73	32.13	31.17	30.20	–
NLM W						
[45, 46]	–	–	31.67	–	29.47	–
NLM SAP						
[14]	34.91	33.23	31.89	30.78	29.83	28.33
NLM PCA						
(PH) [15]	35.39	33.38	32.21	31.07	30.20	28.70
BM3D PCA						
[8]	35.77	34.10	32.91	31.90	31.07	29.75

interesting to see that the best performance was obtained for $h = \sigma_n$ in general for almost all the test images, these results of PSNR are similar or better with respect to those reported in other works in literature (see [37, 43, 45]), also the performance is competitive with those results obtained using the approaches NLM SAP, and NLM PCA overall for some large values of σ_n . For this experiment, one concludes that the best method was BM3D SAPCA, followed by NLM PCA, and NLM SAP, which means that performance of hybrid NLM based methods using patch wise filtering is excellent.

5.2 PSNR results for database TID2008

The results obtained with the second proposed experiment concern the obtention of the PSNR for low level of noise added to a collection or database TID2008 which has been used recently as a good noise-free database with 25 images of very high quality [35]. The main intention for this experiment is to show how much geometry of the noise-free images is removed or preserved by the analyzed approaches. Table 6 shows the PSNR average of the 25 images filtered for the best performing approaches. The classic NLM for $h = \sigma_n$, NLM with Hilbert for $d = 4$ and Tukey kernels with $h = \sqrt{6}\sigma_n$, NLM SAP, NLM PCA (PH), and BM3D SAPCA. From the PSNR results it is clear that the worst approaches for

Table 3 PSNR results in dB for evaluating some NLM methods (Cameraman image)

Method \ σ_n	10	15	20	25	30	40
NLM Buades						
$h = 0.7\sigma_n$	31.89	30.40	29.17	28.19	27.39	25.95
$h = \sigma_n$	31.90	30.64	29.67	28.76	27.87	26.17
$h = 1.5\sigma_n$	31.13	29.82	28.60	27.41	26.29	24.56
NLM h est.						
$h = 1.5h_0$	26.86	26.53	26.13	25.69	25.23	24.27
$h = h_0$	29.67	28.67	27.58	26.59	25.66	23.95
NLM Hilb.						
$d = 2$	28.09	27.45	26.78	26.11	25.47	24.33
$d = 3$	29.82	29.04	28.23	27.39	26.54	25.03
$d = 4$	30.68	29.61	28.52	27.52	26.57	24.97
NLM other						
[22, 42] Tukey	33.18	31.32	30.23	29.58	28.08	26.44
[22, 42] Wave	33.07	31.13	30.06	29.41	27.80	25.85
NLM SAP						
[14]	33.50	31.51	30.32	29.63	28.73	27.97
NLM PCA						
(PH) [15]	33.67	31.29	29.75	28.64	27.70	26.16
BM3D PCA						
[8]	34.48	32.25	30.72	29.61	28.75	27.48

low level of noise are the classic NLM and NLM with Hilbert kernel (loosing geometry and smoothing textures), since for all the noisy cases ($\sigma_n = 1, 3, 5$) it is better do not perform a filtering task. In the case of the Tukey kernel, it is shown that the low level noisy images are well filtered preserving details an geometry, gaining in average 0.25 dB for noise level of $\sigma_n = 1$, 1.68 dB for $\sigma_n = 3$, and 2.79 dB for $\sigma_n = 5$. Equally, for NLM SAP the average gains were 0.45 dB, 1.91 dB, and 3.04 dB respectively, for NLM PCA (PH) average gains were 0.69 dB, 2.18 dB, and 3.57 dB, and finally for BM3D SAPCA the average gains were 1.04 dB, 2.76 dB, and 3.87 dB, giving the best performance, preservation of geometry and textures. Here, it is important to comment that for $\sigma_n = 0$ this last approach fails, which is an interesting drawback provided that for almost all the noisy images its performance is remarkable (the centered line in Table 6, means that any numerical value was obtained, while Inf means an infinite value in dB).

5.3 Results from the subjective point of view

On the other hand, evaluating the subjective aspect of the approaches, Figs. 1 to 9 show some visual results of the denoising task for some benchmark images previously presented in Section 5.1. Figure 1 shows the filtering results of the Barbara image with a level of noise such that $\sigma_n = 20$. From Figs. 2 and 3 one can appreciate better the filtering of two zoomed zones in the scene of Barbara (crops of the full image), the books at the back top-left, and the leg and arm at the bottom-left, in this two figures one compares the best results obtained from all simulated approaches. Some other visual results are also given for the images of

Table 4 PSNR results in dB for evaluating some NLM methods (Mandril image)

Method \ σ_n	10	15	20	25	30	40
NLM Buades						
$h = 0.7\sigma_n$	30.34	28.47	27.01	26.03	25.30	24.42
$h = \sigma_n$	30.36	28.14	26.59	25.18	24.18	22.86
$h = 1.5\sigma_n$	28.87	26.20	24.55	23.49	22.77	21.88
NLM h est.						
$h = 1.5h_0$	26.27	25.61	24.94	24.30	23.72	22.76
$h = h_0$	29.19	27.67	26.47	25.46	24.57	23.13
NLM Hilb.						
$d = 2$	26.15	25.10	24.30	23.66	23.15	22.34
$d = 3$	28.32	26.78	25.60	24.74	24.13	23.21
$d = 4$	29.53	27.66	26.47	25.57	24.84	23.70
NLM other						
[22, 42] Tukey	30.45	28.71	27.29	25.11	23.93	22.55
[22, 42] Wave	29.65	28.25	27.15	25.05	23.95	21.90
NLM SAP						
[14]	30.46	28.69	27.37	26.59	25.12	22.68
NLM PCA						
(PH) [15]	32.27	29.70	27.84	26.50	25.47	23.86
BM3D PCA						
[8]	32.32	29.97	28.17	26.89	26.07	24.54

Cameraman and Boats, for both images the filtering is led by considering AWGN with $\sigma_n = 30$ (see Cameraman in Fig. 4 and Boats in Fig. 7). Also, the resulting denoised images are compared with respect to the original image (free of noise), in some zooming zones in the scenes of the Cameraman and Boats images, the head of the man, arms and the camera in Fig. 5, and the buildings suited at the bottom-right of the camera's tripod in Fig. 6, the lighthouse of the port in Fig. 8 and into the largest boat next of a man where one can see the name of the boat in Fig. 9.

The subjective or visual results obtained using NLM Tukey, NLM PCA and BM3D approaches seems to perform as directional filters, observing Figs. 1, 4, and 7, allowing to a better reconstruction of some details of high frequency into the scenes, where for example, the shadow of the table projected into the arm of Barbara is good restored, also some high frequencies of the pants of Barbara are well recuperated (see Figs. 3c, e, and f), in the case of NLM PCA one can see an over-representation and NLM SAP gives a better smoothed visual result (see sub-figure 3d). The same aspects are present in the image of the bookcase, where some shapes have been best preserved, this is appreciated at the bottom-right side of the bookcase, the books, the objects over the table and some details in the wall (see Figs. 2c, e, and f). Also, for the zooming in Figs. 5, 6, 8 and 9, one can see in the Cameraman image and in the Boats image, that the effects of the reconstruction are more clear for all the obtained results with the different approaches. In the case of the head of the Cameraman one can see that the edges are well preserved with the NLM Tukey, the same symptom is presented for the buildings, the lighthouse of the port and the back's name of the boat. As previously commented, even if the PSNR of the classic NLM is good in general, the recuperated images

Table 5 PSNR results in dB for evaluating some NLM methods (Boats image)

Method \ σ_n	10	15	20	25	30	40
NLM Buades						
$h = 0.7\sigma_n$	33.09	31.39	30.02	28.89	27.95	26.44
$h = \sigma_n$	33.19	31.47	30.14	29.06	28.14	26.71
$h = 1.5\sigma_n$	31.89	29.95	28.44	27.27	26.43	25.20
NLM h est.						
$h = 1.5h_0$	29.45	28.75	28.02	27.31	26.63	25.39
$h = h_0$	31.90	30.14	28.65	27.45	26.39	24.64
NLM Hilb.						
$d = 2$	30.26	29.04	28.07	27.22	26.52	25.39
$d = 3$	32.11	30.67	29.49	28.54	27.72	26.29
$d = 4$	32.77	31.06	29.66	28.53	27.58	26.01
NLM other						
[22, 42] Tukey	33.57	31.53	30.19	28.79	28.01	26.08
[22, 42] Wave	33.37	31.23	29.76	28.20	27.16	25.68
NLM SAP						
[14]	33.49	32.08	30.71	30.04	29.53	28.35
NLM PCA						
(PH) [15]	33.94	31.96	30.65	29.57	28.73	27.34
BM3D PCA						
[8]	34.73	32.78	31.43	30.42	29.15	28.20

are little over smoothed in the case of high frequency objects (see sub-figures 5b, 6b, 8b and 9b), whereas the NLM SAP approach tries to preserve the details but at the same time it smooths the noise components, giving a little over smoothing of the high frequency details

Table 6 Average PSNR results in dB for evaluating some NLM methods for the database TID2008 (25 images), and average computation times of each method

Method \ σ_n	0	1	3	5	Times (sec.)
PSNR (Noise)	Inf	48.15	38.78	34.54	
NLM Buades					
$h = \sigma_n$	Inf	37.94	35.24	34.44	82.72
NLM Hilb.					
$d = 4$	35	33.61	33.25	32.72	84.91
NLM Tukey					
$h = \sqrt{6}\sigma_n$	Inf	48.40	40.46	37.33	80.74
NLM SAP					
[14]	332.81	48.60	40.69	37.58	10.81
NLM PCA					
(PH) [15]	318.71	48.84	40.96	38.11	4.31
BM3D PCA					
[8]	–	49.19	41.54	38.41	123.44

and textures (see sub-figures 5d, 6d, 8d and 9d), particularly for the case of the Fig. 8d the cross over the dome of the lighthouse has been vanished. Finally, the performance of NLM PCA and BM3D SAPCA is remarkable, in the most of situations is over to 1 dB with respect to NLM SAP (over to 2 dB with respect to classic NLM method), but contrary to NLM SAP approach it gives an over representation (rare patch effect) of the high frequencies and models better the textures (see sub-figures 5e, 6e, 8e and 9e), whereas BM3D SAPCA gives the better preservation of edges and texture (see sub-figures 5f, 6f, 8f and 9f). One can say, that obtained results with NLM SAP are between NLM PCA (PH) and BM3D SAPCA, even if in the most of the situations (see sub-figures 5d, 6d, 8d and 9d) it presents some smooth effects due to the ringing effects annulment. In general, BM3D SAPCA approach gives the best results according to the edge preservation and texture modeling, but the computation times were the longest (see Table 6).

In general, one can see that the denoisig is very similar for the case of the three images (the same symptom for objective results obtained in Section 5.2 for lower level of noise). Particularly, the results obtained with the classic NLM and NLM Tukey are in the low range of the state of the art methods (approaches 4, 5, and 6), their performance is satisfactory compared with the excellent performance of the NLM SAP, the NLM PCA (PH) fast algorithm and the BM3D SAPCA. As in the case of the Barbara image, for the Cameraman and Boats images, BM3D SAPCA and NLM PCA permit a good preservation of some high frequency details which one can see in the original images, whereas the obtained results with the classic NLM seems to be a little over smoothed and with a ringing effect. Finally, one can see that for all approaches the reconstruction of some details is really a hard challenge (for high level of noise), for example the cables of the boats found in the original scene of the Boats image are lost in almost all the denoised images, the same reconstruction problem is presented in the Cameraman image, where the textures of the land are almost lost. In all conducted experiments, the size of the search region for approaches 1, 2, and 3 was $\mathcal{I} = [-5, 5] \times [-5, 5]$, and the size of the patch windows or neighborhood was $\mathcal{N}_i = 3 \times 3$. The results obtained for other test images were similar and coherent with respect to those obtained for the Barbara, Cameraman and Boats images.

At this stage, answers to the asked questions in the introduction section could be given. The change of a weighting robust function instead of the exponential function could improve the classic NLM method, this has been shown when using the Tukey and Wave functions. Moreover, the change of the geometry of the searching region and the patch wise filtering ideas have led to approaches that give significant improvements to the classic NLM method. According to some state of the art methods, future works in NLM methods are related to meliorations in the sense of steerable filtering where one can consult recent propositions made by Deledalle [14] and Maleki [31], robust filtering [20], and some others [37]. In particular, we are interested in applications of NLM methods in optics to ameliorate the filtering performance of fringe patterns and ESPI phase-maps [21].

6 Discussion

According to the results obtained in Section 5 from the comparison of six different approaches, and other results that have been obtained in the adequacy of these methods for phase maps filtering, generally NLM filtering can provide results of robustness, competing

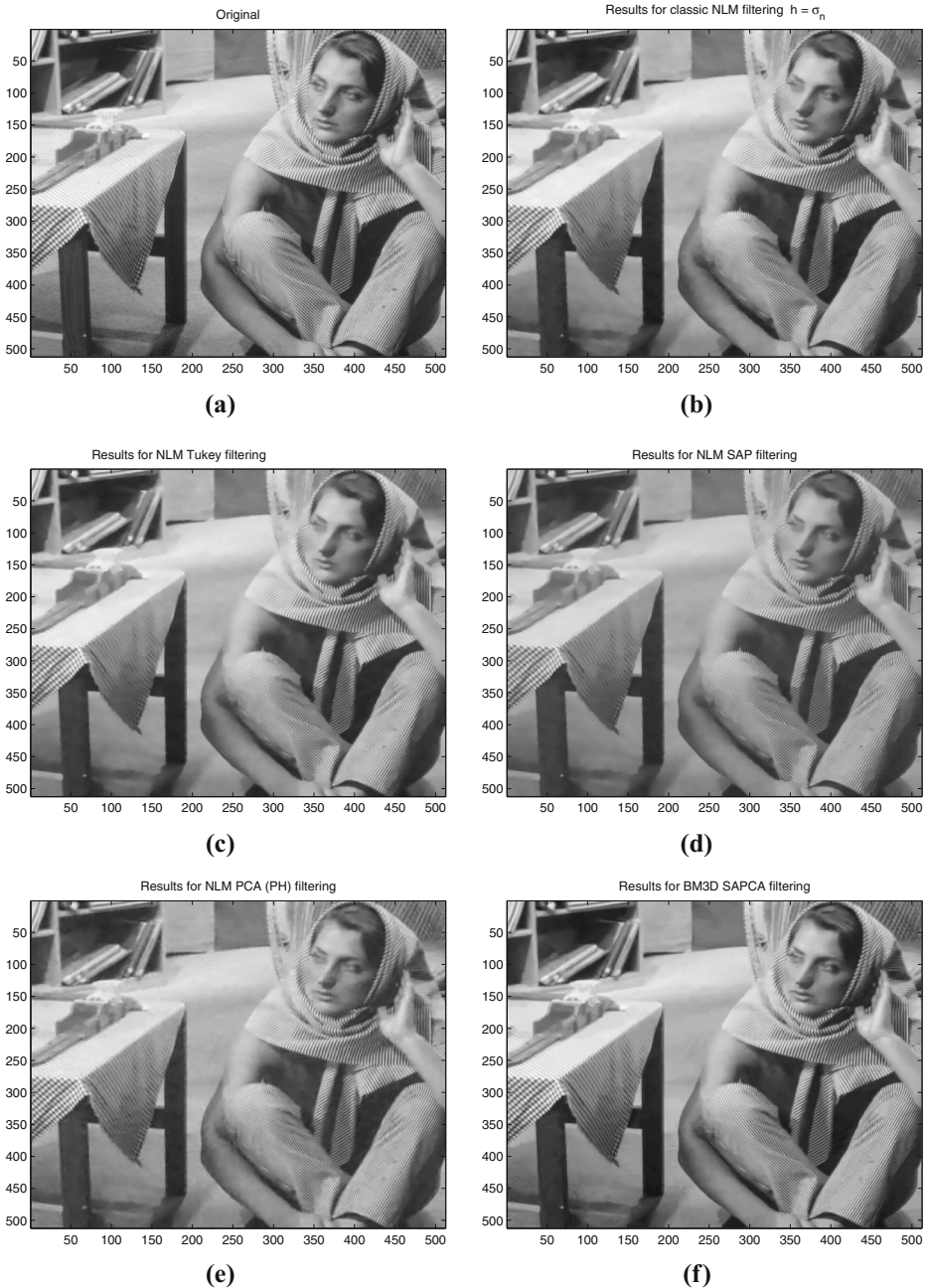


Fig. 1 Results for Barbara as a test image: (a) Barbara image free of noise (Noisy image using Normal pdf with $\sigma_n = 20$), (b) Filtered image using Buades proposition with $h = \sigma_n$, (c) Filtered image using Tukey kernel with $h = \sqrt{6}\sigma_n$, (d) Filtered image using NLM SAP, (e) Filtered image using NLM PCA (PH), (f) Filtered image using BM3D SAPCA

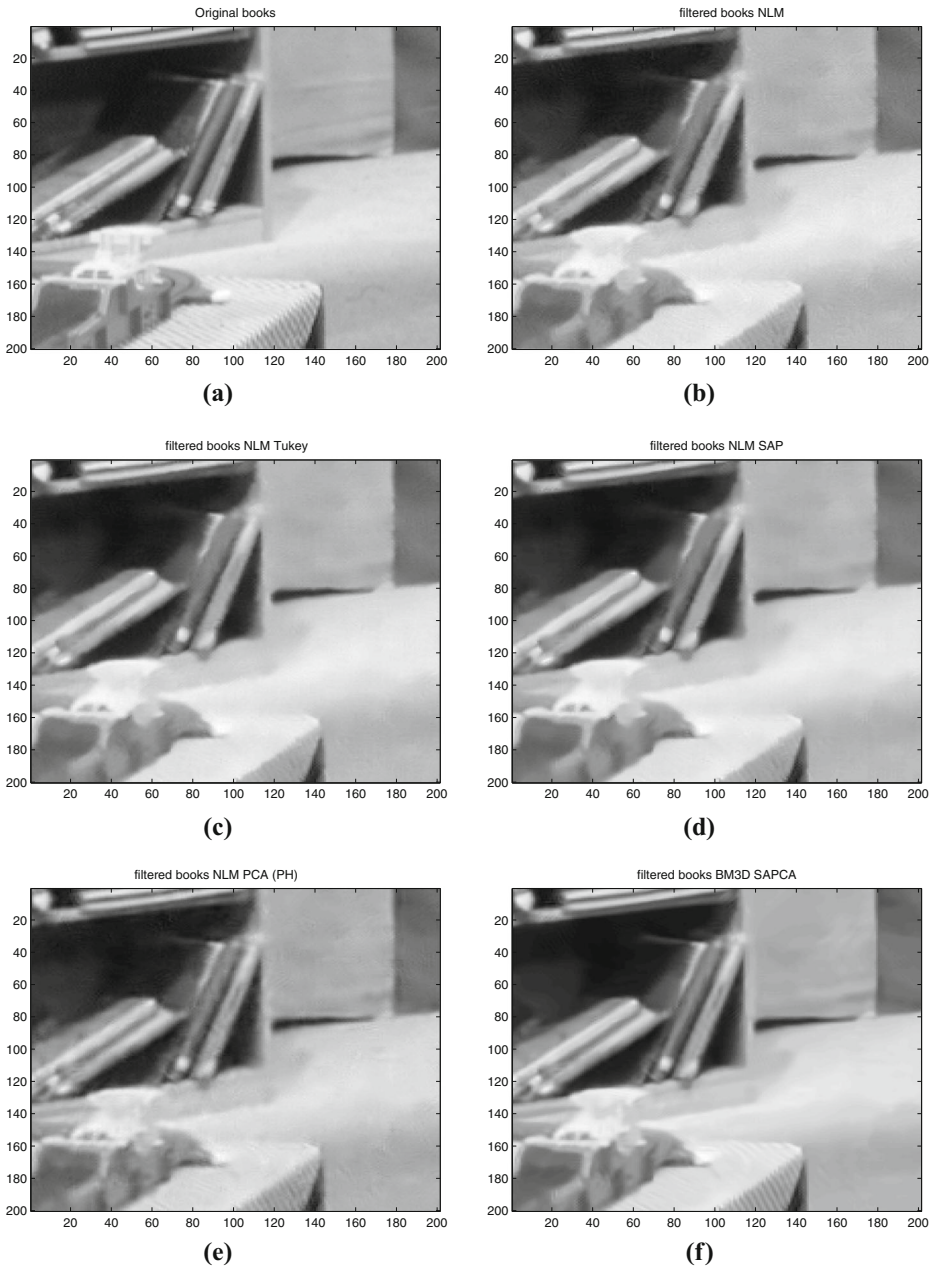


Fig. 2 Results for a zoom of Barbara, books in back: (a) Original books; (b) Filtered books using Buades NLM, (c) Filtered books using Tukey kernel, (d) Filtered books using NLM SAP, (e) Filtered books using NLM PCA (PH), (f) Filtered books using BM3D SAPCA

with other recent filtering methods like directional filtering as the one presented in [21]. The recent hybrid algorithms are based on the change of the kernel, the geometry of the neighborhood and include collaborative filtering stage. The dimensionality reduction of the

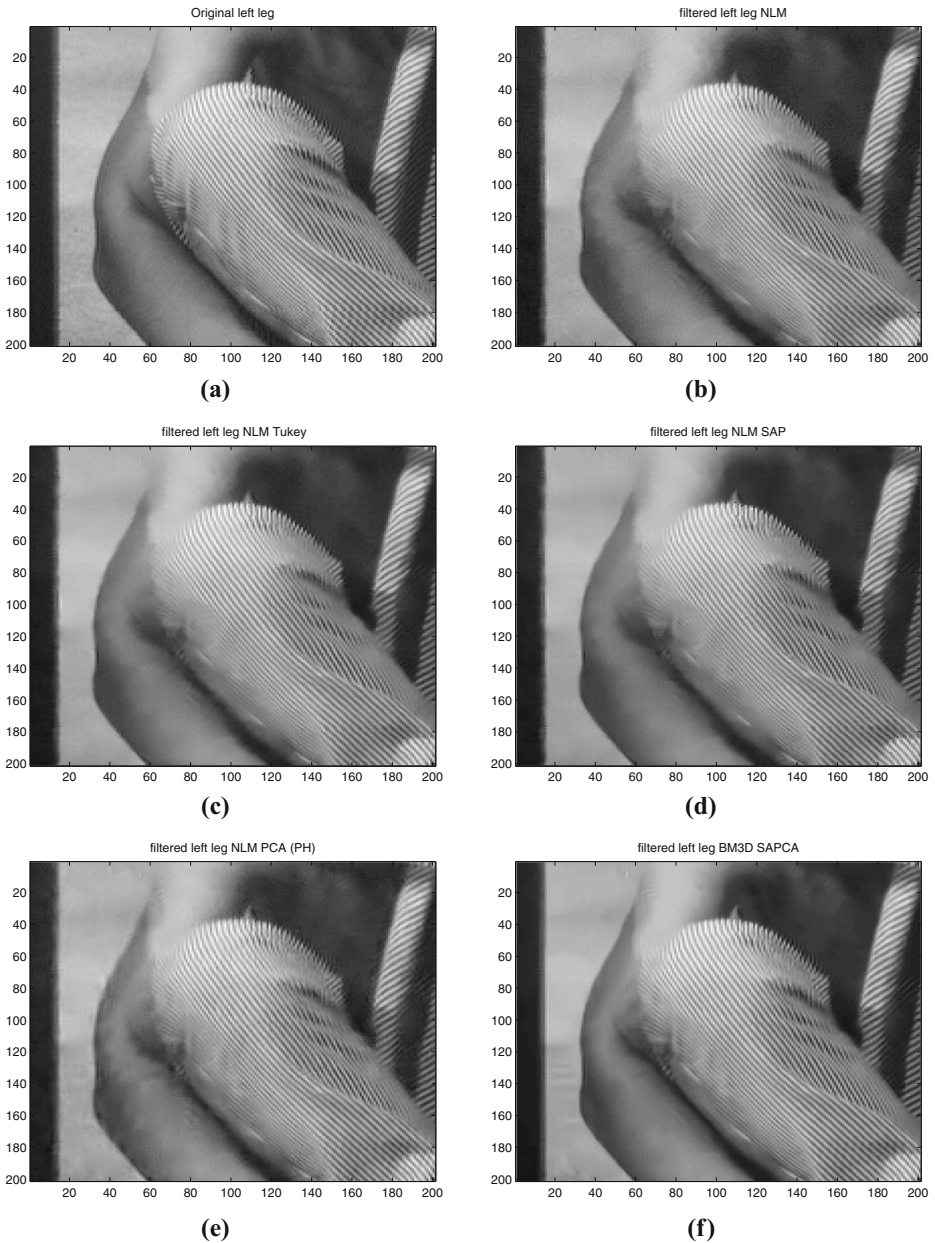


Fig. 3 Results for a zoom of Barbara, left leg in front: (a) Original left leg and part of the arm; (b) Filtered left leg using Buades NLM, (c) Filtered left leg using Tukey kernel, (d) Filtered left leg using NLM SAP, (e) Filtered left leg using NLM PCA (PH), (f) Filtered left leg using BM3D SAPCA

neighborhood or patches onto principal component analysis (PCA) improves the performance of the classic NLM, since the methodology can model far better the geometries and textures of images (NLM SAP, NLM PCA, and BM3D SAPCA).

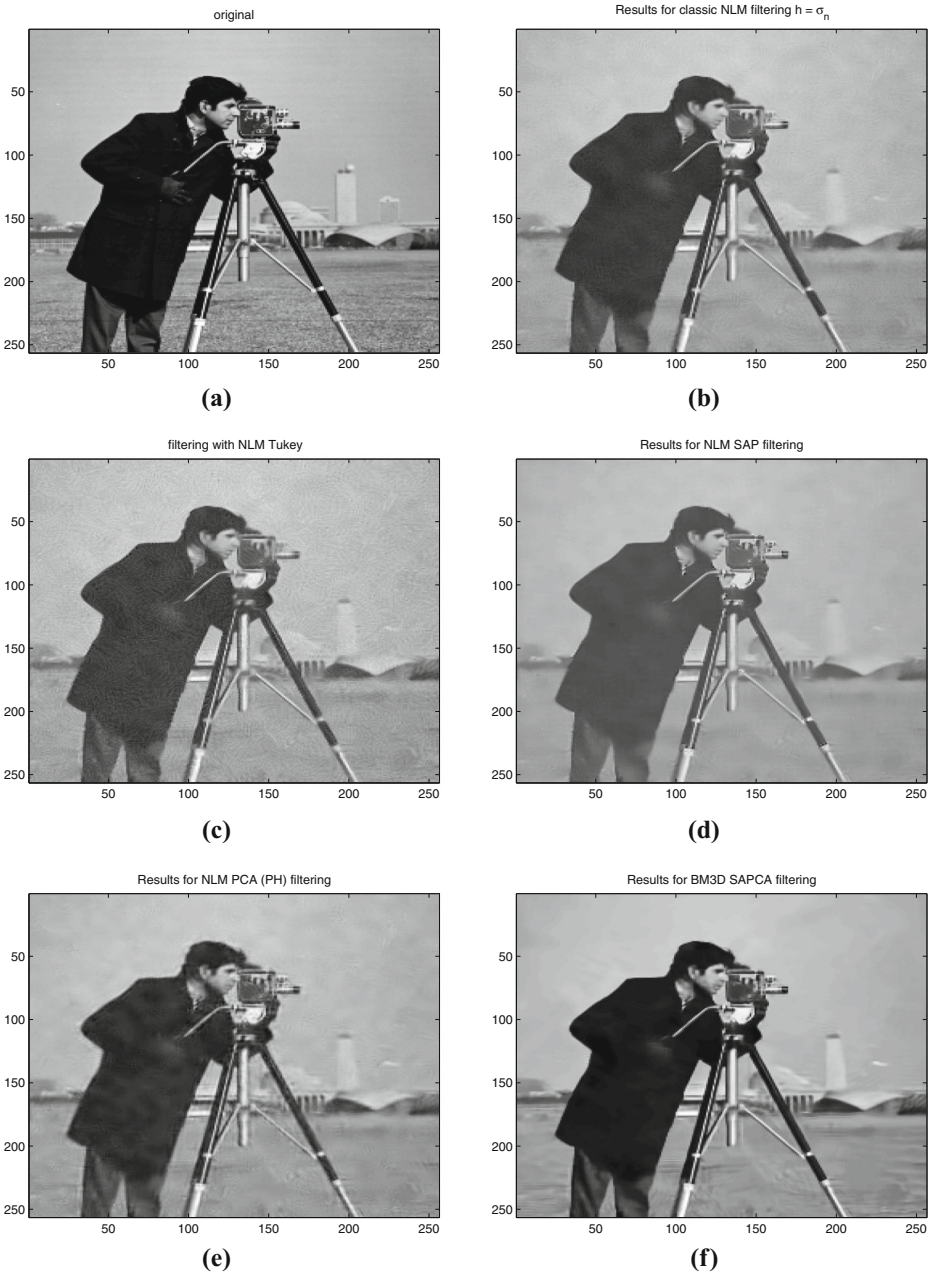


Fig. 4 Results for Cameraman as a test image: (a) Cameraman image free of noise, (b) Filtered image using classic NLM (Noise with $\sigma_n = 30$), (c) Filtered image using Tukey kernel with $h = \sqrt{6}\sigma_n$, (d) Filtered image using NLM SAP, (e) Filtered image using NLM PCA (HP), (f) Filtered image using BM3D SAPCA

On the other hand, in most of the comparison cases it is frequently used the hypothesis of Gaussian noise, since it is easy to generate, and it is always known the standard deviation σ_n used to generate it, which is also a parameter related with the bandwidth h used by NLM

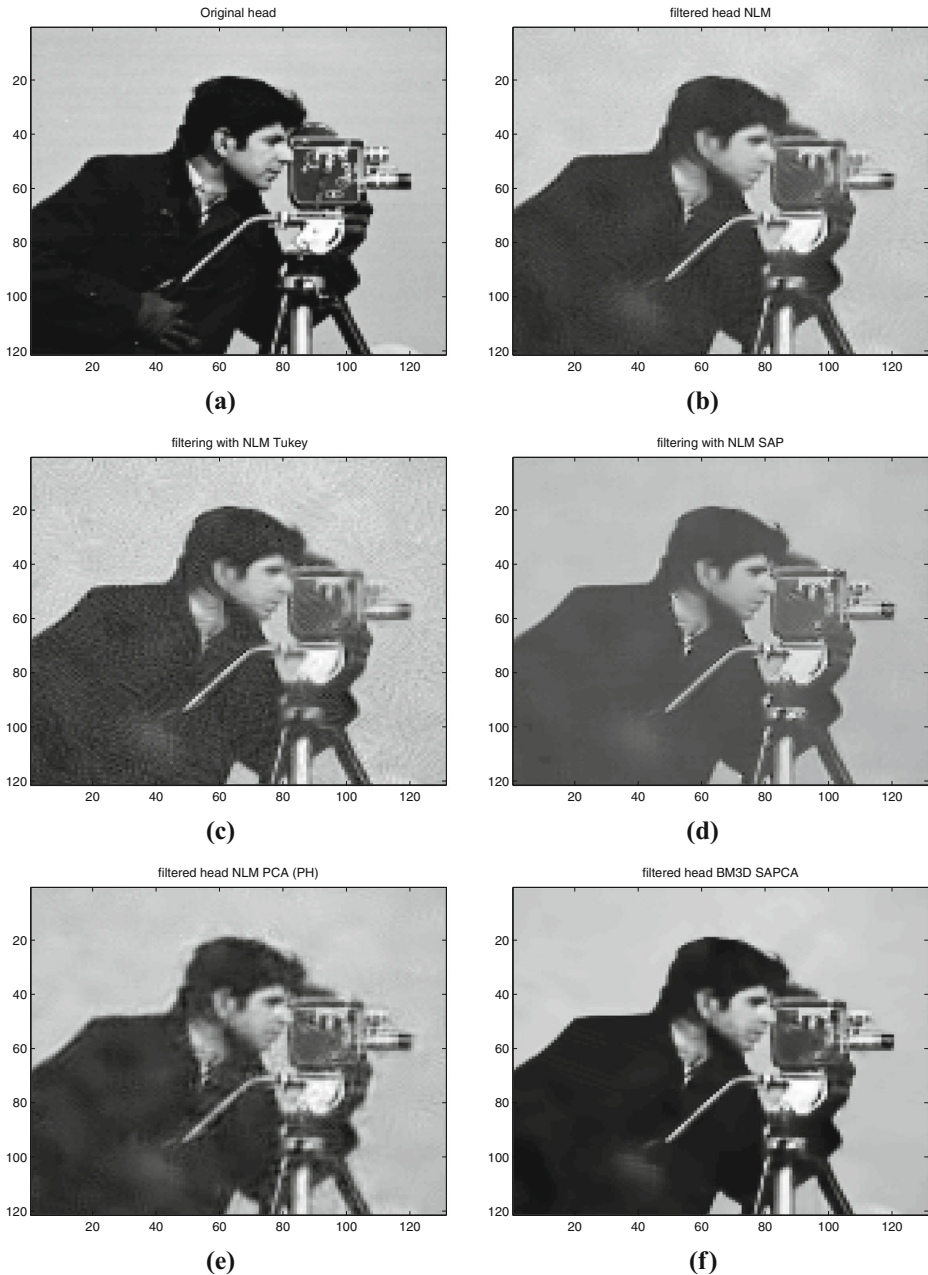


Fig. 5 Results for a zoom of Cameraman, head and arm of man: **(a)** Original head; **(b)** Filtered head using classic NLM, **(c)** Filtered head using Tukey kernel, **(d)** Filtered head using NLM SAP, **(e)** Filtered head using NLM PCA (HP), **(f)** Filtered head using BM3D SAPCA

filtering ($h = k_0\sigma_n$). However, in practice σ_n is always unknown, or the noise distribution is non Gaussian in some applications and then it is essential to estimate the noise level or a noise variance function as proposed in [6, 9, 35, 36]. For example, NLM has been adapted

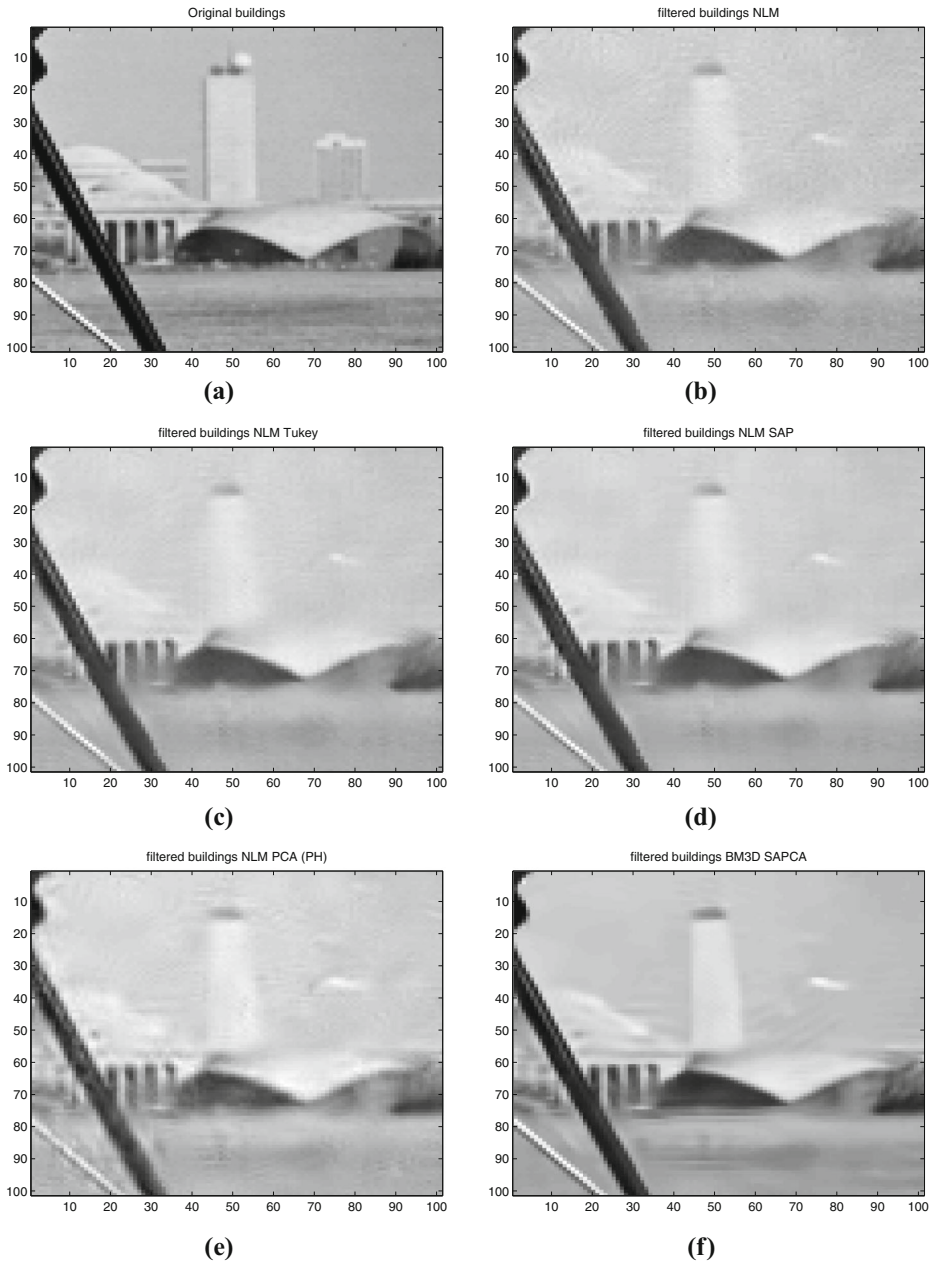


Fig. 6 Results for a zoom of Cameraman, buildings in bottom right: (a) Original buildings; (b) Filtered buildings using classic NLM, (c) Filtered buildings using Tukey kernel, (d) Filtered buildings using NLM SAP, (e) Filtered buildings using NLM PCA (HP), (f) Filtered buildings using BM3D SAPCA

for fringe and ESPI phase-map denoising, where the noise is considered of speckle type; in this case we could not adapt the BM3D SAPCA for the ESPI phase-map denoising, and also comparison is made with respect of other proposed method (Continuous Wavelet Transform – CWT) introduced in [21].

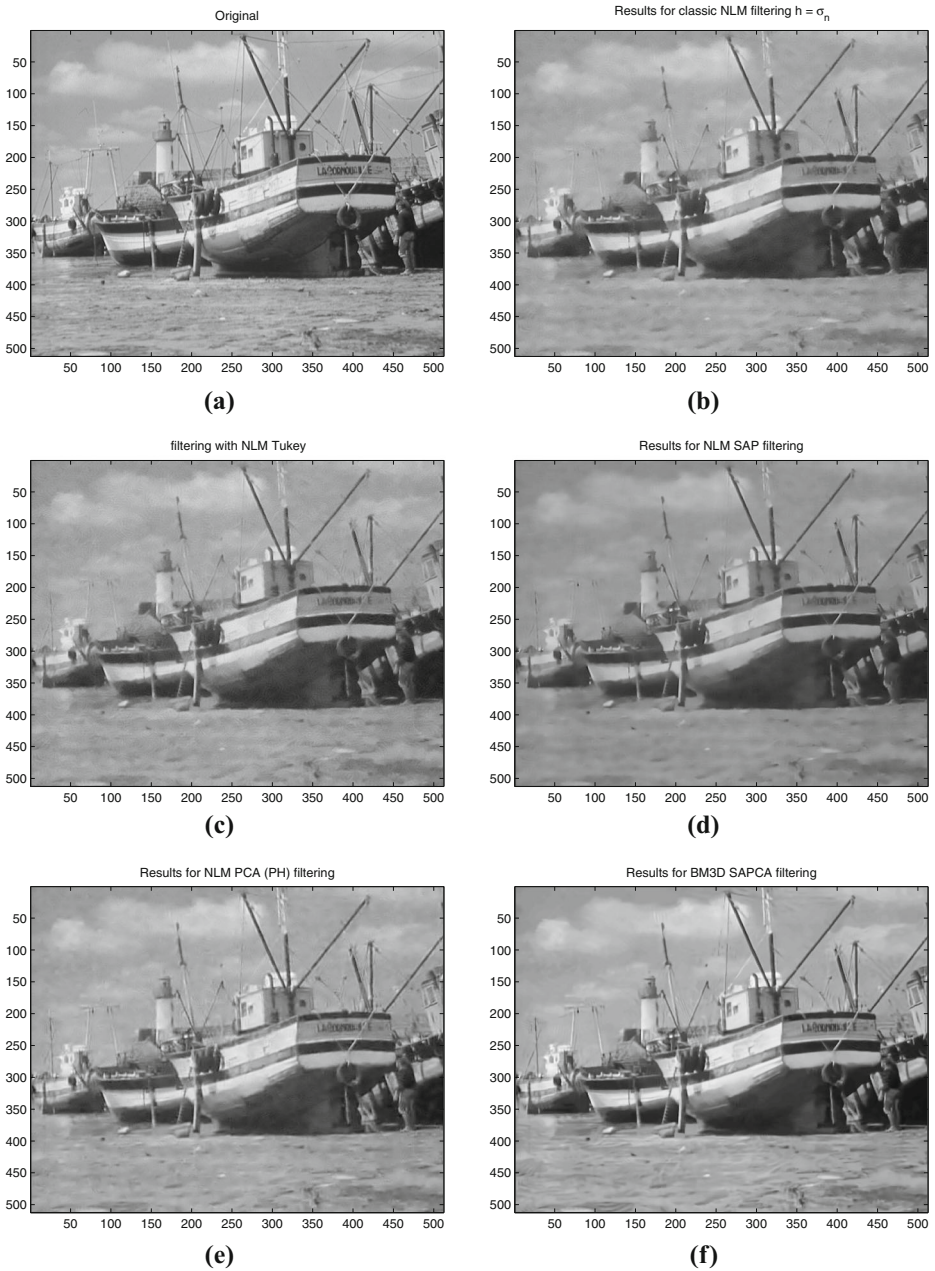


Fig. 7 Results for Boat as a test image: (a) Boat image free of noise, (b) Filtered image using classic NLM (Noise with $\sigma_n = 30$), (c) Filtered image using Tukey kernel with $h = \sqrt{6}\sigma_n$, (d) Filtered image using NLM SAP, (e) Filtered image using NLM PCA (HP), (f) Filtered image using BM3D SAPCA

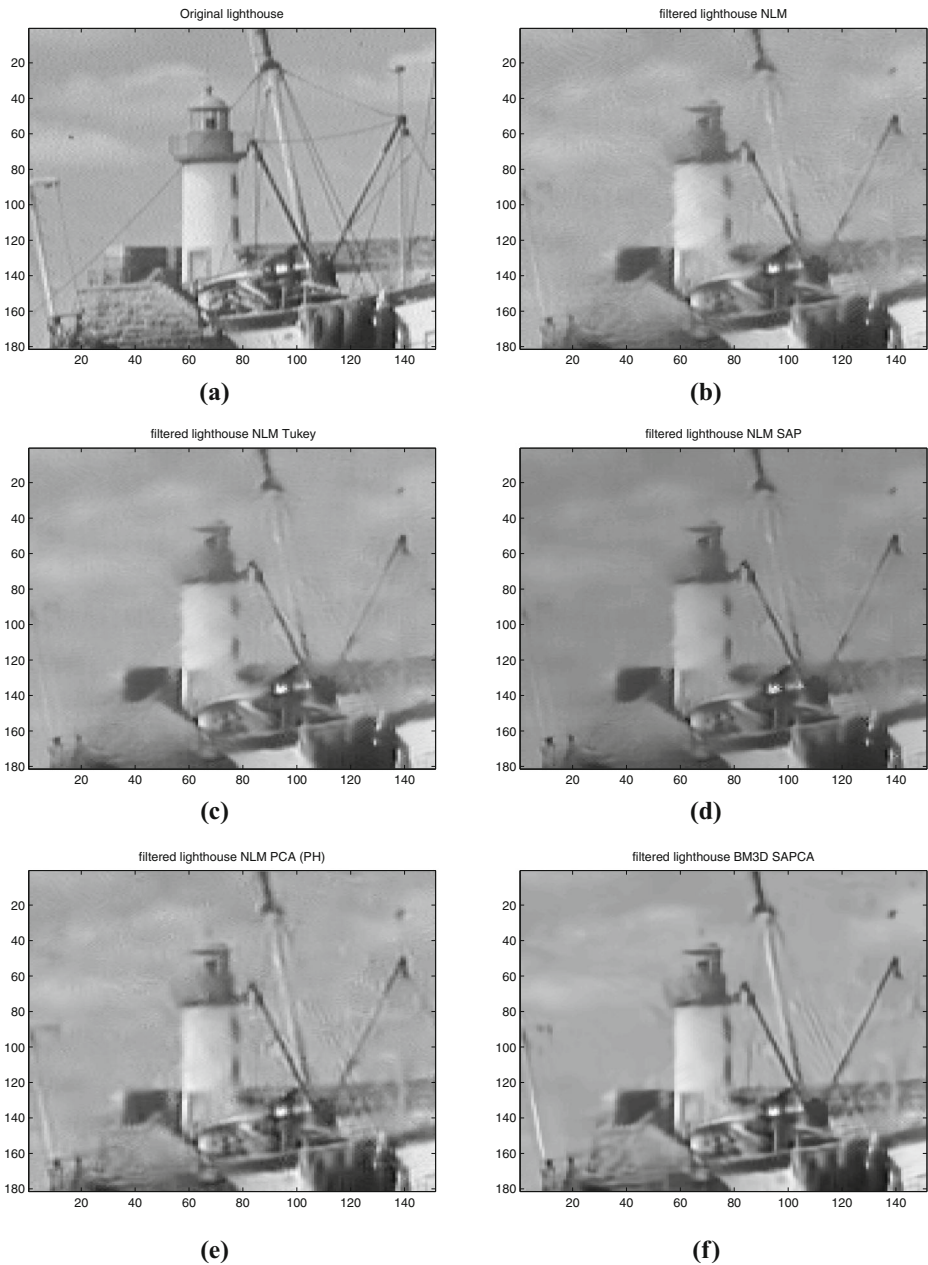


Fig. 8 Results for a zoom of Boat, lighthouse port: (a) Original lighthouse; (b) Filtered lighthouse using classic NLM, (c) Filtered lighthouse using Tukey kernel, (d) Filtered lighthouse using NLM SAP, (e) Filtered lighthouse using NLM PCA (HP), (f) Filtered lighthouse using BM3D SAPCA

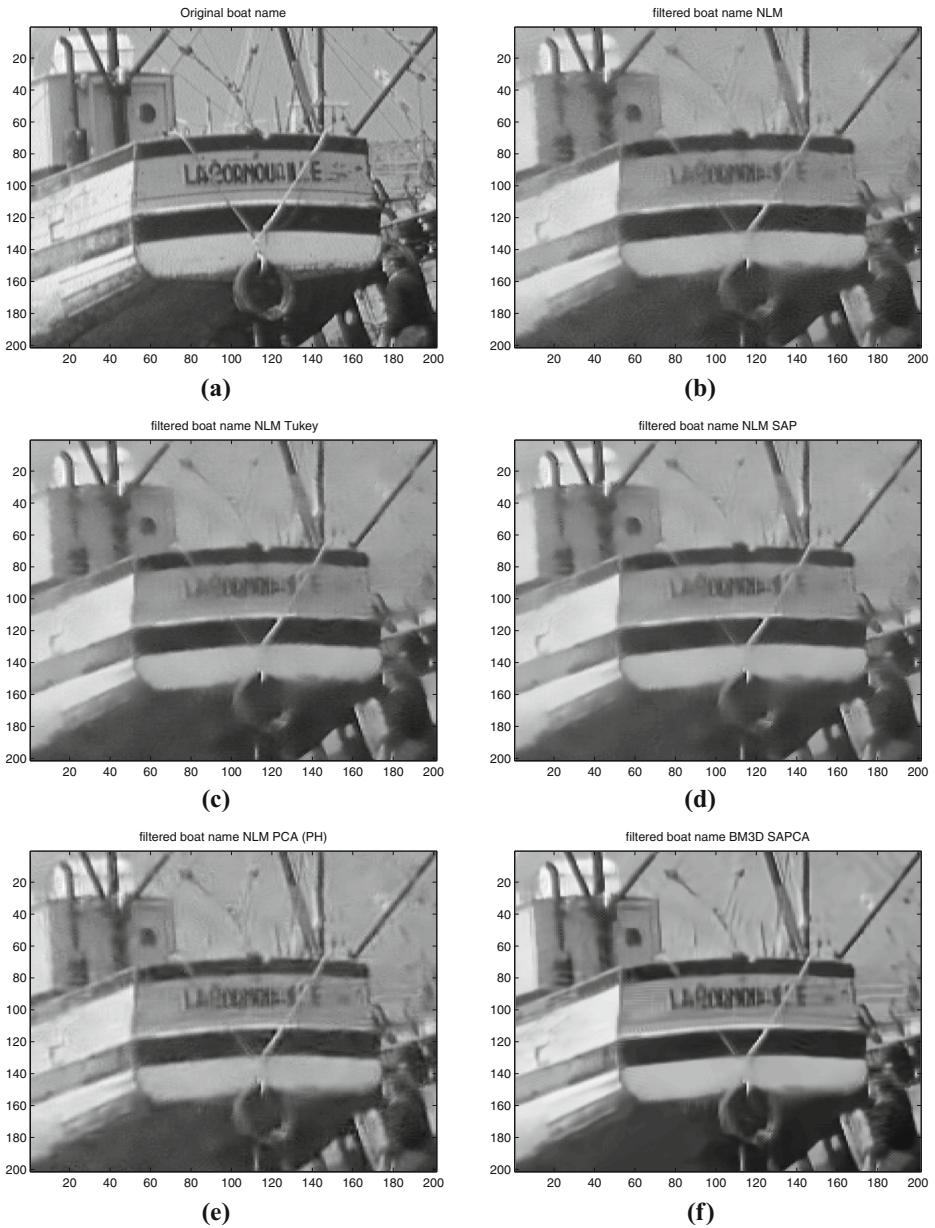


Fig. 9 Results for a zoom of Boat, the name and back of the largest boat: (a) Original boat name; (b) Filtered name using classic NLM, (c) Filtered name using Tukey kernel, (d) Filtered name using NLM SAP, (e) Filtered name using NLM PCA (HP), (f) Filtered name using BM3D SAPCA

7 Conclusions

A review and comparison of different NLM approaches for digital image filtering have been conducted in the present paper. The performance of classic NLM filtering could be improved when changing the kernel, the use of three different kernels has been analyzed (Hilbert, Tukey, and Wave). Particularly, it has been shown that for the case of the NLM Tukey kernel it gives a better denoising results than classic NLM. On the other hand, the excellent performance given by recent hybrid approaches such as NLM SAP, NLM PCA (PH), and the BM3D SAPCA lead to establish that significantly improvements to the classic NLM could be obtained. Particularly, the BM3D SAPCA approach gives the best denoising results, however, the computation time was the longest. According to this last case, if one is restricted by the time of computation the NLM SAP and NLM PCA (PH) denoising are good options to be selected depending on the type of images. If the images are rich on textures NLM PCA (PH) is recommended, if images have more edges and homogeneous regions than textures the NLM SAP is then recommended.

Acknowledgments The first author acknowledge all support given by the Universidad Autónoma de Zacatecas (UAZ) during the years 2014–2015 to realize one academic year (invited professor-researcher) of leave at the Instituto Tecnológico de Sonora (ITSON). Additional acknowledgements for the support given by CONACyT, particularly the postgraduate programs in Electrical Engineering Sciences of both institutions (UAZ and ITSON). Also, our deep gratitude to the Secretaría de Educación Pública (SEP) of Mexico, this work was partially supported through the PIFI 2014–2015 Mexican Program from UAZ. Finally, we acknowledge the valuable comments and recommendations made by two anonymous reviewers.

References

1. Berlinet A, Devroye L (1994) A comparison of kernel density estimates. *Publications de l'Institut de Statistique de l'Université, de Paris* 38(3):3–59
2. Buades A, Coll B, Morel JM (2005) A non-local algorithm for image denoising. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol 2, San Diego, USA, pp 60–65
3. Buades A, Coll B, Morel JM (2010) Image denoising methods. A new Nonlocal principle. *SIAM Rev* 52(1):113–147
4. Buades A, Coll B, Morel JM (2008) Nonlocal image and movie denoising. *Int J Comput Vision* 76(2):123–139
5. Buades A, Coll B, Morel JM (2011) Self-similarity-based image denoising. *Commun ACM* 54(5):109–117
6. Colom M, Buades A, Morel J-M (2014) Nonparametric noise estimation method for raw images. *J Opt Soc Am A* 31(4):863–871
7. Dabov K, Foi A, Katkovnik V, Egiazarian K (2007) Image denoising by sparse 3-D Transform-Domain collaborative filtering. *IEEE Trans on Image Processing* 16(8):2080–2095
8. Dabov K, Foi A, Katkovnik V, Egiazarian K (2009) BM3D Image denoising with Shape-Adaptive Principal Component Analysis. In: *Proceedings of the Workshop on Signal Processing with Adaptive Sparse Structured Rep. (SPARS'09)*, Saint-Malo, France
9. Danielyan A, Foi A (2009) Noise variance estimation in nonlocal transform domain. In: *Proceedings of the International Workshop on Local and Non-Local Approx. in Image Proc. (NLNA2009)*, Tuusula, Finland, pp 41–45
10. De la Rosa JI, Fleury G (2002) On the Kernel selection for Minimum-Entropy estimation. In: *Proceedings of the IEEE Instrumentation and Measurement Technology Conference*, Anchorage, AK (USA), pp 1205–1210
11. De la Rosa JI, Fleury G, Davoust M-E (2003) Minimum-entropy, pdf approximation and Kernel selection for measurement estimation. *IEEE Trans Instrum Meas* 52(4):1009–1020

12. De la Rosa JJ, Villa JJ, De la Rosa E, González E, Gutiérrez O, Escalante N, Ivanov R, Fleury G (2013) MAP Entropy estimation: Applications in robust image filtering. *Journal of the European Optical Society-Rapid Publication* 8:13047
13. Deledalle C-A, Denis L, Tupin F (2009) Iterative weighted Maximum Likelihood denoising with probabilistic patch-based weights. *IEEE Trans on Image Processing* 18(12):2661–2672
14. Deledalle C-A, Duval V, Salmon J (2012) Non-local Methods with Shape-Adaptive Patches (NLM-SAP). *J Math Imaging Vision* 43(2):103–120
15. Deledalle C-A, Salmon J, Dalalyan AS (2011) Image denoising with patch based PCA: Local versus global. In: *Proceedings of the British Machine Vision Conference (BMVC)*. University of Dundee
16. Devroye L (1992) A note on the usefulness of superkernels in density estimation. *Ann Stat* 20:2037–2056
17. Devroye L (1989) The double kernel method in density estimation. *Ann Inst Henri Poincare* 25:533–580
18. Devroye L (1997) Universal smoothing factor selection in density estimation: theory and practice. *Test* 6:223–320
19. Devroye L, Krzyżak A (1999) On the Hilbert kernel density estimate. *Statistics and Probability Letters* 44:299–308
20. Dinesh PJ, Govindan VK, Abraham TM (2009) Robust estimation approach for nonlocal-means denoising based on structurally similar patches. *International Journal of Open Problems in Computer Science and Mathematics* 2:293–310
21. Escalante N, Villa J, De la Rosa I, De la Rosa E, González-Ramírez E, Gutiérrez O, Olvera C, Araiza M (2013) 2-D Continuous Wavelet transform for ESPI phase-maps denoising. *Opt Lasers Eng* 51(9):1060–1065
22. Goossens B, Luong H, Pizurica A, Philips W (2008) An improved non-local denoising algorithm. In: *Proceedings of International Workshop on Local and Non-Local Approximation in Image Processing (NLNA2008)*, Lausanne, Switzerland, pp 143–156
23. <http://www.cs.tut.fi/~foi/GFC-BM3D/>
24. <http://www.math.u-bordeaux1.fr/~cdeledal/>
25. <http://www.mathworks.com/matlabcentral/fileexchange/13176-non-local-means-filter>
26. Juditsky A, Nemirovsky A (2010) Nonparametric denoising signals of unknown local structure, II: Nonparametric function recovery. *Appl Comput Harmon Anal* 29:354–367
27. Lebrun M, Buades A, Morel J-M (2013) Implementation of the Non-local Bayes (NL-bayes) image denoising algorithm. *Image Processing On Line (IPOL)*:1–42
28. Lin L, Lingfu K (2009) Image denoising base on Non-local means with Wiener filtering in wavelet domain. In: *Proceedings of the Fifth IEEE International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pp 471–474
29. Liu YL, Wang J, Chen X, Guo Y-W, Peng Q-S (2008) A robust and fast non-local means algorithm for image denoising. *J Comput Sci Technol* 23(2):270–279
30. Loader CM (1999) Bandwidth selection: classical or plug-in? *Ann Stat* 27(3):415–438
31. Maleki A, Narayan M, Baraniuk RG (2013) Anisotropic nonlocal means denoising. *Appl Comput Harmon Anal* 35:452–482
32. Masry E (1983) Probability density estimation from sampled data. *IEEE Trans Inf Theory* 29(5):697–709
33. Milanfar P (2013) A tour of modern image filtering. *IEEE Signal Process Mag* 30(1):106–128
34. Ponomarenko NN, Lukin VV, Zriakhov MS, Kaarna A, Astola J (2007) An automatic approach to lossy compression of AVIRIS images. In: *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium-IGARSS*, pp 472–475
35. Pyatykh S, Hesser J, Zheng L (2013) Image noise level estimation by principal component analysis. *IEEE Trans on Image Processing* 22(2):687–699
36. Sutour C, Aujol J-F, Deledalle C-A (2016) Automatic estimation of the noise level function for adaptive blind denoising. In: *Proceedings of the 24th IEEE European Signal Processing Conference (EUSIPCO)*, pp 76–80
37. Sutour C, Deledalle CA, Aujol JF (2014) Adaptive regularization of the NL-means: Application to image and video denoising. *IEEE Trans on Image Processing* 23(8):3506–3521
38. Takeda H, Farsiu S, Milanfar P (2007) Kernel regression for image processing and reconstruction. *IEEE Trans on Image Processing* 16(2):349–366
39. Talebi H, Zhu X, Milanfar P (2013) How to STAF-ly boost denoising performance. *IEEE Trans on Image Processing* 22(4):1470–1485
40. Terrell GP (1990) The maximal smoothing principle in density estimation. *J Am Stat Assoc* 85:470–477
41. Terrell GP, Scott DW (1985) Oversmoothed nonparametric density estimation. *J Am Stat Assoc* 80:209–214

42. Tian J, Yu W-Y, Xie S-L (2008) On the kernel function selection of nonlocal filtering for image denoising. In: Proceedings of the IEEE Seventh International Conference on Machine Learning and Cybernetics, pp 2964–2969
43. Van De Ville D, Kocher M (2009) SURE-Based non-Local means. *IEEE Signal Process Lett* 16(11):973–976
44. Van De Ville D, Kocher M (2011) Non-local means with dimensionality reduction and SURE-based parameter selection. *IEEE Trans on Image Processing* 20(9):2683–2690
45. You SJ, Cho NI (2011) A new image denoising method based on the wavelet domain Nonlocal means filtering. In: Proceedings of the ICASSP, vol 2011, pp 1141–1144
46. You SJ, Cho NI (2013) An adaptive bandwidth nonlocal means image denoising in wavelet domain. *EURASIP Journal on Image and Video Processing* 60:1–22



José I. de la Rosa was born in Zacatecas, Mexico, on March 19, 1972. He received the B.Sc. degree in electronics and telecommunications engineering from the Universidad Autónoma de Zacatecas (UAZ), Mexico, in 1995, the M.Sc. degree in digital systems from the Instituto Politécnico Nacional (CITEDI-IPN), Mexico, in 1998, and the Ph.D. degree in signal processing and control from the Université de Paris-Sud, Orsay, France, in November 2002. He was a Ph.D. student within the Measurement Department of SUPELEC. Presently, he is a Full time Professor-Researcher at the engineering faculty from the UAZ, Mexico. His main interests are statistical signal processing, modeling and system identification, and speech signal and image processing. Using MCMC, Markov random Fields, Directional Transforms, and Non parametric estimation for Inverse Problems.



Jesús Villa-Hernández received the B.Sc. degree in electronics and communications from the Universidad Autónoma de Zacatecas (UAZ), México, in 1993, and the Ph.D. degree from the Centro de Investigaciones en Óptica (CIO), León, Guanajuato, Mexico, in 1999. He was a postdoctoral student in the Facultad de Ciencias Físicas, Universidad Complutense de Madrid, Spain, in 2002. His research interests are digital image processing, optical metrology and computer vision. He is currently a titular researcher in the Unidad Académica de Ingeniería Eléctrica, UAZ.



Joaquín Cortez received the B.Sc. degree in electronics in 1994 from the Instituto Tecnológico de Sonora (ITSON). In 2000 he earned the M.Sc. degree in electrical engineering focusing on telecommunications engineering from the Centro de Investigación y Estudios Avanzados (CINVESTAV) - IPN, Guadalajara, Mexico. In 2008 he also obtained the Ph.D. degree in electrical engineering focusing on telecommunications engineering from the Centro de Investigación y Estudios Avanzados del IPN, Guadalajara, Mexico. Presently, he is a Full time Professor-Researcher at the electrical and electronics engineering department from ITSON. His main interests are statistical signal processing in communications, MIMO systems, and Space temporal Coding.



Hamurabi Gamboa received his Bachelor in Electronics and Communications Engineering in the Faculty of Engineering of the University of Guadalajara in 2000. From 2001 to 2003, He completed his Master degree in Electrical Engineering from, focusing on the Digital Signal Processing at the University of Guanajuato. He completed his doctoral studies at the Technical University of Dresden, Germany in the area of Voice Processing 2010. Currently works as a professor and researcher at the Academic Unit of Electrical Engineering of the Autonomous University of Zacatecas, Mexico in the area of research of digital signal processing.



José G. Arceo received the B.Sc. degree in electronics and telecommunications engineering from the Universidad Autónoma de Zacatecas (UAZ), Mexico, in 2001. Obtained the M.Sc. degree in digital systems from the Instituto Politécnico Nacional (CITEDI-IPN), Mexico, in Tijuana, Baja California, in 2003. He completed the Ph.D. focusing in telecommunications and electronics from the Escuela Superior de Ingeniería Mecánica y Eléctrica (ESIME), IPN, in January 2008. His main interests are statistical signal processing in communications, digital image processing, and optical metrology applications.



Efrén González received his B.Sc. in electronics engineering from the University of Guanajuato, Salamanca, Guanajuato, Mexico, working in Image Processing, “Border Detection in Aerial Images”, in March 1999. The M.Eng. was also earned at the University of Guanajuato, Salamanca, Guanajuato, Mexico in June 2000. Specialized in Signal Processing, “Model of the Initial Position of a GPS with Selective Availability Present”. His Ph.D. degree was earned in July 2005 at Bristol University, Bristol, United Kingdom. Weather Radars, “Weather Radar Data Analysis Oriented to Improve the Quality of Rainfall Estimation”. Presently, he is Professor-Researcher in DSP at the Autonomous University of Zacatecas, Zacatecas, Mexico from January 2007. His main interests are on weather radar, Adaptable signal processing, and image processing topics, and teaching activities for engineering undergraduate and postgraduate students.