MEDICAL IMAGE DENOISING USING KERNEL RIDGE REGRESSION

Dinh Hoan TRINH\textsuperscript{1}, Marie LUONG\textsuperscript{2}, Jean-Marie ROCCHISANI\textsuperscript{3}, Canh Duong PHAM\textsuperscript{4}, Françoise DIBOS\textsuperscript{1}

\textsuperscript{1} LAGA, Université Paris 13, 99 Avenue Jean-Baptiste Clément, Villetaneuse, France
\textsuperscript{2} L2TI, Université Paris 13, 99 Avenue Jean-Baptiste Clément, Villetaneuse, France
\textsuperscript{3} Hôpital Avicenne - Médecine Nucléaire, Bobigny, France
\textsuperscript{4} CIID, Vietnam Academy of Science and Technology, Hanoi, Vietnam

ABSTRACT

Medical images are often corrupted by random noise, leading to undesirable visual quality. Thus, image denoising is one of the fundamental tasks required by medical imaging analysis. In this paper, we propose a novel learning method for the reduction of Gaussian noise of Computed Tomography (CT) image and Rician noise of Magnetic Resonance Imaging (MRI) image based on a given set of standard images and the Kernel Ridge Regression (KRR). Experimental results demonstrate the outperformance of the proposed technique over various other methods in terms of both objective and subjective evaluations.

Index Terms— Nonlinear regression, Medical image denoising, Kernel Ridge Regression, CT image, MRI image.

1. INTRODUCTION

Medical images acquired from CT and MRI instruments are often affected by random noise, resulting in a loss in image quality and a reduction of the visibility of image features especially in low contrast regions. Such effects can thereby compromise the accuracy and the reliability of pathological diagnosis or surgery purposes. One of the required preprocessing steps for medical application is hence noise removal techniques which consist to determine the best estimate of the original image from its noisy one while preserving edges of relevant structures. The classical denoising technique such as the Gaussian Filter (GF) \cite{1} is efficient for smooth region but limited by blurring effects in high activity regions such as edges and textures. In order to overcome this drawback, many edge preserving filters have been proposed, including the Anisotropic Diffusion Filter (ADF) \cite{2}, the Total Variation (TV) \cite{3,4,5}, and more recently the Non-local means (NL-means) \cite{6}. The ADF attempts to preserve edges by convolving the image in the orthogonal direction of the local gradient. Although straight edges can be well preserved, curved edges or features are usually degraded. Likewise, the total variation minimization based noise removal approach can effectively preserve straight edges. However, details can be over smoothed, depending on the compromise between noise removing and edges preserving (value of the Lagrange multiplier parameter). In the popular NL-means \cite{6} method, a denoised value of a pixel is estimated by the weighted average value of the most resembling pixels which have no reason to be close. However, since the image is noisy, it is clear that using a weighted average of all pixels in the noisy image to recover the original image is not guaranteed. In fact, some of important small details can also be lost. However, the most important problem in the medical image denoising is how to preserve small details as much as possible, as subtle details could raise up essential diagnosis information.

It is well known in medical imaging that many images were acquired at approximately the same location. Thus, using a set of standard (acceptably and proven by experts as noise-free) images to denoise a new noisy image is very useful. In this paper, we introduce a new edge-preserving denoising method for CT and MRI image, that produces better image quality than available alternatives. The basic idea of the method is to formulate the image denoising problem as a nonlinear KRR using training set which is established from a set of given standard images. In order to adaptively remove noise, we first propose to classify the training set into subsets according to some features, namely homogeneous zones, edges/Textures zones and luminance. Then, corresponding KRR function for each subset is computed. Afterwards, these functions are used for denoising selectively, according to the feature of pixel in noisy images. Experimental results show that our method (referred as KRRD) not only effectively removes noise but also well preserves small details.

The rest of this paper is organized as follows. In Section 2, we briefly review the KRR technique. Section 3 describes our proposed algorithm. Our experiments and their results are reported in Section 4. Section 5 concludes the article.

2. KERNEL RIDGE REGRESSION

Suppose that we have a training set \{\((x_1, y_1), \ldots, (x_n, y_n)\)\} with vectors \(x_i \in \mathbb{R}^d\) and associated response value \(y_i \in \mathbb{R}, i = 1, \ldots, n\) (given by a supervisor). The goal of the regression problem is to estimate a function \(f\) which mini-
mizes some measure of discrepancy between the estimation 
\( y_i = f(x_i) \) and the value of \( y_i \). KRR is a well-known statistical technique to solve non linear regression and often outperforms other techniques. Its basic idea relies on mapping the data into a higher dimensional space \( H \) (also called feature space) according to a mapping \( \Phi \), and then finding a linear regression function with the new training set \( \{(\Phi(x_1), y_1), \ldots, (\Phi(x_n), y_n)\} \), which represents a nonlinear regression in the original input space. Let us begin with the linear ridge regression problem in \( H \) [7].

Without loss of generality we can assume that the training set has zero-mean. The linear ridge regression problem consists in minimizing the following cost:

\[
F(w) = \sum_{i=1}^{n}(y_i - \langle w, \Phi(x_i) \rangle)^2 + \lambda\|w\|^2
\]  

(1)

where \( \lambda \) is a fixed positive number used as regularization parameter to control the trade-off between the bias and variance of the estimate. In [7], it is shown that the predicted label \( y \) (i.e. \( y = \langle w, \Phi(x) \rangle \)) of a new unlabeled sample \( x \) is:

\[
f(x) = Y^T(K + \lambda \mathbf{I})^{-1}\kappa
\]  

(2)

where \( K = [(\Phi(x_1), \Phi(x_2))_{\times n}, \kappa = (\langle \Phi(x_1), \Phi(x) \rangle, \ldots, (\Phi(x_n), \Phi(x)))^T \) and \( Y \) is the vector of values \( y_i \). It is hence easy to generalize the ridge regression to KRR by using kernel trick, with a mapping \( \Phi \) such that kernel function \( k : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R} \) represents the scalar product:

\[
\langle \Phi(x_1), \Phi(x_2) \rangle = k(x_1, x_2).
\]

(3)

The equation (2) is then the kernel ridge regression predictor which can be rewritten in the following form:

\[
f(x) = a_1k(x_1, x) + a_2k(x_2, x) + \ldots + a_nk(x_n, x)
\]  

(4)

where \( (a_1, a_2, \ldots, a_n) = Y^T(K + \lambda \mathbf{I})^{-1} \). In this work, the following radial basic function is chosen as kernel function:

\[
k(x, x) = \text{exp}(-||x_j - x||^2/h^2)
\]  

(5)

where \( h \) is the decay parameter. Note that \( \Phi \) need not be made explicit as long as we can obtain the kernel function, which in turn should satisfy Mercer’s condition [9].

3. NOISE-INDUCED KRR

In this section, we apply the KRR to the medical image denoising problem, with the following assumptions:

i) Noisy image \( X \) (to denoise) is corrupted by additive noise \( N(0, \sigma) \) with zero-mean and standard deviation \( \sigma \):

\[
X = H + N(0, \sigma).
\]  

(6)

where \( H \) is the ideal image. It has been established that the noises on CT and MRI images were found to have a Gaussian distribution [10] and Rician distribution [11], respectively.

ii) We have standard CT (MRI) images \( \{A_m\} \) which are acceptably proven by experts as noise-free images and at approximately the same location of \( X \) in the body.

The proposed method includes essentially the training phase and the denoising phase. In the training phase:

1) For a given noise level (e.g. estimated noise level of noisy image), establish the training set from a given set of standard images by adding the noise to the standard images.

2) Perform a robust classification of the training set into groups according to some image features, namely homogeneous zones, texture/edge zones and luminance.

3) Determine for each training group, the corresponding nonlinear KRR function.

Note that, training phase can be performed off-line. In the denoising phase, estimation of each pixel is performed using the KRR function which is computed with the same noise level. These phases are detailed in the next subsections.

3.1. Training Phase

In this phase, training set generation and classification are performed followed by the computation of KRR.

In order to establish the training set \( G = \{(x_1, y_1), \ldots, (x_n, y_n)\} \) from a given set of standard images \( \{A_m\} \), we first create a set of noisy images \( \{B_m\} \) as follows:

\[
B_m = A_m + \tilde{N}(0, \tilde{\sigma})
\]  

(7)

where \( \tilde{N}(0, \tilde{\sigma}) \) is the noise with standard deviation \( \tilde{\sigma} \) that can be estimated from \( X \). In this paper, we use method in [12] to estimate standard deviation of Gaussian noise on CT image, while Rician noise on MRI image can be estimated based on method in [13]. Then, for each pixel \( i \) of \( B_m \), a pair \( (x_i, y_i) \) is determined where \( x_i \in \mathbb{R}^d \) is the vector corresponding to a square patch of fixed size \( d = (2s+1) \times (2s+1) \) and centered at pixel \( i \), while \( y_i \) is the value of pixel \( i \) of \( A_m \) (see Fig. 1).

In order to adaptively estimate the value of a noisy input, samples of similar characteristics should be used to construct the KRR function, which will be applied for denoising this input. So, the training set is classified into \( K \) groups according to the characteristic vectors defined by features, namely homogeneous zone, texture/edge zone and luminance. Here, training set classification is achieved by first applying the SVD to the gradient field of each patch (see [14]) and then using \( k \)-means clustering. Unlike existing works, we propose to classify each patch based on the magnitude of the singular values and its mean value, by defining for each sample \( x_i \) in
the training set, the characteristic vector $v_i$ as follows:

$$v_i = (\lambda_1^i - \lambda_2^i, \mu_i) \in \mathbb{R}^2$$

(8)

where $\lambda_1^i, \lambda_2^i$ are singular values of the patch corresponding to vector $x_i, \mu_i$ is the mean of pixel values in $x_i$. For a patch in a smooth region, there is no dominant direction and all eigenvalues are small. In the case of oriented edge/texture region, there is a dominant direction and the corresponding eigenvalue is larger than the others.

The classification of the training set into $K$ groups is performed by classifying the characteristic vectors into $K$ clusters according to the $k$-means clustering technique. $K$ can be achieved using histogram of standard images. The $k$-means clustering aims to partition the $n$ vectors $v_1, \ldots, v_n$ into $K$ ($2 \leq K < n$) clusters $\Omega = \{\Omega_1, \ldots, \Omega_K\}$ while minimizing the following optimization problem:

$$\arg\min_{\Omega} \sum_{i=1}^{K} \sum_{v_j \in \Omega_i} \|v_j - \nu_i\|^2$$

(9)

where $\nu_i$ is the mean of elements in $\Omega_i$. After performing the classification, we obtain the training set $G = \bigcup_{k=1}^{K} G_k$, where $G_k = \{(x_i, y_i) \in G | v_i \in \Omega_k\}$ with characteristic vector $\nu_i \in \mathbb{R}^2$.

Finally, functions $F_1^{KRR}, \ldots, F_K^{KRR}$ corresponding to groups $G_1, \ldots, G_K$ can be computed (section 2).

3.2. Image Denoising

Denoising an image $X$ (6) is performed by using the KRR functions that are computed with the estimated noise level of $X$. For each pixel corresponding to input vector $x$, denoised value is adaptively computed by using the KRR function of the group which $x$ belongs to. In other words, this estimated value only depends on the training group which has similar characteristics to $x$. Note that the sizes of $X$ and standard images can be different. The algorithm is as follows.

**Step 1:** For each of pixel $i$ in image $X$, first establish a vector $x \in \mathbb{R}^d$ corresponding to a patch size of $(2s + 1) \times (2s + 1)$ and centered at $i$, then compute its characteristic vector $v = (\lambda_1 - \lambda_2, \mu)$ as in (8). Note that size of $x$ and size of samples $x_i$ in the training set are the same.

**Step 2:** Determine the Group $G_k$ that $x$ belongs to:

$$k = \min_i \{\|v - \nu_i\|^2, i = 1, 2, \ldots, K\}.$$  

(10)

**Step 3:** Use KRR function $F_k^{KRR}(\cdot)$ of the group $G_k$ to estimate the value of pixel $i$.

4. EXPERIMENTAL RESULTS

Several experiences have been carried out on CT and MRI images. Here, only five examples of test CT and MRI images are reported (Fig. 3). For each test image, a training set is established by using three standard images (not necessarily identical to the test image). Fig. 2 only illustrates one standard image for each example. Three cases are considered when the test image is corrupted by additive Gaussian noise (and Rician noise) with zero-mean and $\sigma = 20, 25$ and 30. We use patch size of $3 \times 3$ and $h$ in (5) is set to $1.5\hat{\sigma}$. We use cross-validation [8] to find $\lambda$ in (1). A comparative evaluation using objective quality measures has been performed to demonstrate the advantages of our method (KRRD) over other denoising methods, namely, the GF, the ADF, the TV minimization and the NL-means proposed by [1], [2], [4] and [6] respectively. Two objective metrics, namely the PSNR and the SSIM [15] are used to evaluate the fidelity. The PSNR measures the intensity difference between two images. However, it is well-known that it can fail to describe the subjective quality of the image. The SSIM is one of the most frequently used metrics for image quality assessment. Compared with the PSNR, it better expresses the structure similarity between the recovered image and the reference one. The results are reported in Table 1 and Table 2. As can be seen, our method yields significant PSNR and SSIM gains over the other methods. Due to the lack of space, only results for $\sigma = 30$ are reported in Table 2. Denoised images are also reported for subjective comparison. As shown in the results of the TV and NL-means methods (Fig. 4 and Fig. 5), certain subtle details are lost compared to our method where noise is effectively removed in the smooth region while better preserving edges around the object contours. In particular, TV’s result involves seriously affected quality in smooth region (Fig. 4c). Moreover, residual images of CT images are shown (Fig. 4) to illustrate the outperformance of our denoising method, as our residual image is close to the Gaussian noise and contains nearly no structural element.

5. CONCLUSION

We have proposed a novel denoising method for CT (MRI) images based on KRR and training using standard images. Experimental results demonstrated the superior performances of the proposed method over some well known techniques. We believe that with an effective training set, this technique may be quite useful and promising. Moreover, it can be used for denoising of medical images with known probability den-
Table 1. PSNR comparison of denoised images

<table>
<thead>
<tr>
<th>Test Image</th>
<th>( \sigma )</th>
<th>( \hat{\sigma} )</th>
<th>\text{PSNR (dB)}</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>( \hat{\sigma} )</td>
<td>( \text{GF} )</td>
</tr>
<tr>
<td>(a)</td>
<td>20</td>
<td>18.97</td>
<td>22.265</td>
</tr>
<tr>
<td>(c)</td>
<td>30</td>
<td>27.12</td>
<td>19.406</td>
</tr>
<tr>
<td>(d)</td>
<td>20</td>
<td>18.97</td>
<td>21.612</td>
</tr>
<tr>
<td>(e)</td>
<td>25</td>
<td>22.93</td>
<td>19.057</td>
</tr>
<tr>
<td>(f)</td>
<td>30</td>
<td>27.12</td>
<td>18.894</td>
</tr>
<tr>
<td>(g)</td>
<td>20</td>
<td>18.97</td>
<td>21.612</td>
</tr>
<tr>
<td>(h)</td>
<td>25</td>
<td>22.93</td>
<td>19.057</td>
</tr>
<tr>
<td>(i)</td>
<td>30</td>
<td>27.12</td>
<td>18.894</td>
</tr>
</tbody>
</table>

Table 2. SSIM comparison (\( \sigma = 30 \)) of denoised images

<table>
<thead>
<tr>
<th>Test Image</th>
<th>( \sigma )</th>
<th>( \hat{\sigma} )</th>
<th>\text{SSIM}</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \hat{\sigma} )</td>
<td>( \text{GF} )</td>
<td>( \text{ADF} )</td>
</tr>
<tr>
<td>(a)</td>
<td>20</td>
<td>27.12</td>
<td>0.7648</td>
</tr>
<tr>
<td>(b)</td>
<td>25</td>
<td>27.69</td>
<td>0.8007</td>
</tr>
<tr>
<td>(c)</td>
<td>30</td>
<td>27.62</td>
<td>0.7916</td>
</tr>
<tr>
<td>(d)</td>
<td>20</td>
<td>27.12</td>
<td>0.7648</td>
</tr>
<tr>
<td>(e)</td>
<td>25</td>
<td>27.69</td>
<td>0.8007</td>
</tr>
<tr>
<td>(f)</td>
<td>30</td>
<td>27.62</td>
<td>0.7916</td>
</tr>
</tbody>
</table>

6. REFERENCES