Image Denoising Using Local Adaptive Least Squares Support Vector Regression

WU Dingxue  PENG Daiqiang  TIAN Jinwen

Abstract  Rather than attempting to separate signal from noise in the spatial domain, it is often advantageous to work in a transform domain. Building on previous work, a novel denoising method based on local adaptive least squares support vector regression is proposed. Investigation on real images contaminated by Gaussian noise has demonstrated that the proposed method can achieve an acceptable trade off between the noise removal and smoothing of the edges and details.

Keywords  least square support vector machines; image denoising

CLC number  P237.4; TP753

Introduction

When images are acquired and processed, an “original” desired image may become degraded in various ways. Suppose a signal \( s(i, j) \) is polluted by an additive Gaussian noise \( n(i, j) \) and noisy observation \( y(i, j) \) is represented by the equation \( y(i, j) = s(i, j) + n(i, j) \). The goal of denoising is to remove the noise \( n(i, j) \), while retaining as much as possible the important signal features. Very commonly, this can be achieved by approaches such as Wiener filtering, which is widely used for image denoising. Generally, Wiener theory consists in finding a linear structure of the original spatial domain in the sense of minimizing the mean-square error. This tends to lose some detailed information, and noises in the vicinity of edges are not suppressed well. To overcome this problem, the Wiener filter has been extended to multiple-bases representations for noise removal. Mihcak and Kozintsev\(^{[1]}\) approached the signal estimation problem from the perspective of designing the Wiener filter in the wavelet domain. The technology indirectly yields an estimate of the signal subspace that is leveraged into the design of the filter. This paper studies the problem of nonlinear Wiener filtering in reproducing kernel Hilbert spaces via least square support vector regression. The method reflected new perspectives within the framework of kernel methods for denoising problem. Experimental results confirm a significant improvement in image denoising.

1  Least squares support vector regression

Least support squares vector regression is a new universal learning machine proposed by Suykens et al.\(^{[2]}\). Let \( x \in R^d \), \( y \in R \), \( R^d \) represent input space, \( d \)
is the dimension. By some nonlinear mapping \( \phi \), \( x \) is mapping into some a prior chosen Hilbert space spanned by the linear combination of a set of functions.

\[
y = f(x, \omega) = \sum_{l=1}^{N} \alpha_{l} \phi_{l}(x) + b + e_{i}
\]

(1)

with \( \phi(x) : R^{d} \rightarrow R \).

Such that the following regularized risk function \( J \) is minimized:

\[
J = \frac{w}{2} \sum_{i=1}^{N} e_{i}^{2} + \frac{\gamma}{2} \sum_{i=1}^{N} \alpha_{i}^{2}
\]

(2)

The parameter \( \gamma \) is a positive regularization constant. After elimination of \( w, e \) one obtains the solution:

\[
\begin{bmatrix}
0 & \rho^{-1} \\
\rho & \Omega
\end{bmatrix} \begin{bmatrix}
b \\
\alpha
\end{bmatrix} = \begin{bmatrix}
0 \\
Y
\end{bmatrix}
\]

(3)

where \( Y = [y_{1}, \cdots, y_{N}] \), \( \rho = [1 \cdots 1] \), \( \alpha = [\alpha_{1}, \cdots, \alpha_{N}] \) and \( \Omega = K + \gamma^{-1} I \). The resulting least support squares vector regression model for function estimation becomes:

\[
f(x) = \sum_{i=1}^{N} \alpha_{i} K(x, x_{i}) + b
\]

(4)

where \( K(x, x_{i}) = \phi(x) \phi(x_{i}) (i = 1, \cdots, N) \) is the kernel function and must satisfy the Mercer condition\(^{[3]}\), \( \alpha \) are Lagrange multipliers and \( b \) almost equals the mean of \( y \).

2 Proposed theory and algorithm

Consider a 2D image consisting of a matrix of \( M = N \times N \) pixels, the observation image can be regarded as a function in pixel areas \( y = f(i, j) : R^{2} \rightarrow R^{1} \), where input \((i, j)\) is 2D vector equals to the row and column indices of that pixel, where output \( y \) is the approximated intensity value\(^{[4]}\). The Lagrange multipliers \( \alpha_{i,j} \) of the observed image pixel \( y(i,j) \) can be easily calculated using Eq.(3).

\[
\alpha = A(I - IB)Y = O_{a}Y = O_{a}(S + N)
\]

(5)

where \( A = \Omega^{-1} \), \( B = I^{T} \Omega^{-1} I^{T} \) and \( O_{a} \) is a \( N \times N \) matrix defined by \( A(I - IB) \). Notice that, the Lagrange multipliers \( \alpha_{i,j} \) of the observed image pixel \( y(i,j) \) is determined by the multiplication of the matrix \( O_{a} \) and the observed image \( Y \). That is, the Lagrange multipliers are influenced by the clean image \( S \) and random noise \( N \). As in Eq.(4), the observed image can be reconstructed by a linear combination of kernels with weights equal to the values of Lagrange multipliers and an appropriate support vector regression can concentrate the signal energy into a number of support vectors (SVs) that \( \alpha_{i,j} \) is nonzero. The localization of SVs is particularly appropriate for imaging applications, where it is crucial to preserve fine details like edges and textures. Pixels with positive Lagrange values try to raise the grey levels of themselves and their neighbors, while those with negative Lagrange multipliers will try to reduce the grey levels and they appear darker. Therefore, the Lagrange multipliers effectively weigh the kernel functions to estimate intensity value of image. Furthermore, random noise can be considered as forces that try to make Lagrange multipliers to oscillate above and below the standard value. The noise can be reduced by smoothing the value of Lagrange multipliers, whereas sharp edges may be preserved within certain ranges which rely on a suitable kernel function possessing the capability of nonlinear representation.

With \( \alpha \) in hand, we can attenuate the contribution of a particular kernel function to the signal expansion by weighting the corresponding coefficient \( \alpha_{i,j} \) by a number \( 0 \leq h_{i,j} \leq 1 \). That is, we modify \( \alpha_{i,j} \) component-wise according to \( \hat{\alpha}_{i,j} = h_{i,j} \alpha_{i,j} \) in the sense of minimizing the mean square error. Clearly, the crucial issue is the design of the filter \( h \). Setting \( h_{i,j} = 0 \) completely removes the contribution of kernel function and setting \( h_{i,j} = 1 \) leaves it unaltered. Choosing \( 0 < h_{i,j} < 1 \) attenuates the contribution of corresponding kernel function. Without loss of generality, let \( S \) be a small window of a fixed size \((2R+1)\times(2R+1)\) and the mean of noise to be zero, the resulting local wiener estimator is given by a scalar processor of the form.

\[
\hat{\alpha}_{i,j} = \mu_{a_{i,j}} + (\alpha_{i,j} - \mu_{a_{i,j}}) \max(0, \frac{\sigma_{a_{i,j}}^{2} - \sigma_{n_{i,j}}^{2}}{\sigma_{a_{i,j}}^{2}})
\]

(6)

where \( \sigma_{\cdot,\mu} \) are the local variances and local means in the moving window \( S \), respectively. It can be calculated as follows:

\[
\mu_{a_{i,j}} = \frac{1}{(2R+1)^{2}} \sum_{k=-R}^{R} \sum_{l=-R}^{R} \alpha_{i+k,j+l}
\]

(7)

\[
\sigma_{a_{i,j}} = \frac{1}{(2R+1)^{2}} \sum_{k=-R}^{R} \sum_{l=-R}^{R} (\alpha_{i+k,j+l} - \mu_{a_{i,j}})^{2}
\]

(8)
Finally, we invert the modified Lagrange multipliers to obtain the signal estimate using Eq.(4):

$$\hat{S} = \hat{\alpha}K + b$$

(9)

3 Experimental results and discussion

To verify the proposed algorithm, we present two series of experiments that compare the performance of the proposed approach with that of local Wiener filter in spatial domain. In these experiments, the radial basis function (RBF) is used as the kernel function of least support squares vector regression. For RBF kernel one has:

$$K(x_i,x_j) = \exp\left\{-\|x-x_i\|^2/\sigma^2\right\}$$

(10)

Firstly, we begin with a test function to investigate the edge preserving capability of the proposed denoising algorithm. Filtered results appear in Fig.1. The chosen test function has some abrupt changing features such as discontinuities or sharp bumps. We corrupted the test function by addition simulated spatially Gaussian noise with mean 0 and standard deviation 0.02 shown in Fig.1(a). Fig.1(b) shows the filtered results by local Wiener filter in spatial domain and Fig.1(c) shows the result of the proposed approach. As it appears, Fig.1(b) shows that step edges are poorly filtered by local Wiener filter in spatial domain and the ability of the proposed approach to smooth homogeneous regions simultaneously preserving edges is better than local Wiener filter in spatial domain. Secondly, the standard 256×256 Lena image is presented for additive Gaussian noise. We corrupted it by addition Gaussian noise at different noise levels. The proposed approach is also compared to other techniques. For quantitative evaluation, we use PSNR (peak signal-to-noise ratio) to evaluate the performance of denoising algorithm. Table 1 lists the results of the PSNR improvement of the proposed algorithm compared with other methods-standard LS-SVM, median filter and wiener filter. It is evident from Table 1 that the proposed method consistently outperforms standard LS-SVM and median filter method in terms of PSNR improvement. Fig.2 shows the comparison of the denoised Lena image corrupted with Gaussian noise at PSNR=22.09 dB. As it appears, Fig.2(b) shows that step edges are poorly filtered by standard LS-SVM filter. The median filter and Wiener filter can reduce noise respectively in certain range. Fig.2(e) shows the ability of our approach to smooth homogeneous regions simultaneously preserving edges. Especially, details of the image are very obvious, e.g. eyes and hat. This indicates that the proposed denoising method achieved the most successful noise reduced in statistically homogeneous areas and preserved interesting edge details.

![Fig.1 Investigating edge preserving capability of denoising algorithm via test function](image1)

![Fig.2 Comparison of denoised Lena image corrupted with Gaussian noise at PSNR=22.09 dB](image2)
Table 1  SNRs for Lena image (mean=123.5, std.=43.6) which contain different noising degree by the two denoising methods

| Noisy image | Standard LS-SVM | Median filter | Wiener filter | Proposed method |
|-------------|-----------------|---------------|---------------|-----------------|
| 28.18       | 28.14           | 27.31         | 29.90         | 32.51           |
| 24.61       | 27.63           | 26.89         | 28.93         | 30.89           |
| 22.09       | 26.94           | 26.37         | 27.78         | 29.05           |
| 20.24       | 26.30           | 25.85         | 27.33         | 27.50           |
| 18.69       | 25.64           | 25.37         | 26.67         | 26.65           |

4 Conclusions

We highlighted an efficient approach of nonlinear Wiener filtering that relies on the theory of least support squares vector regression. With the proposed approach, the clean signal and random noise are first mapped from input space into a high-dimensional feature space by a suitable nonlinear mapping $\phi$ and then local Wiener filter is considered in this feature space. This technology provides an elegant way of dealing with the problem of nonlinear Wiener filter by reducing them to linear ones in some feature space nonlinearly related to input space. Furthermore, the number of kernel function satisfying the Mercer condition is enormous. One can also use the wavelet kernel or other kernels as kernel functions to remove noise. Thus, the freedom of the present new registration is very large. Experimental results show that the algorithm can preserve edges with better visual quality and yield better results.

References

[1] Mihcak M K, Kozintsev I, Ramchandran K(1999) Spatially adaptive statistical modeling of wavelet image coefficients and its application to denoising [C]. IEEE International Conference on Acoustics, Speech and Signal Processing, Phoenix

[2] Suykens J A K, Vandewalle J (1999) Least squares support vector machine classifiers [J]. Neural Processing Letters, 9(3):293-300

[3] Vapnik V (1995) The nature of statistical learning theory [M]. New York: Springer-Verlag

[4] Cheng Hui, Tian Jinwen, Liu Jian, et al.(2004) Wavelet domain image denoising via support vector regression [J]. Electron Lett, 40(23):1479-1481

Notes to Contributors

Contributions are welcomed on one of the following subjects or in related areas:

- GIS
- GPS
- RS
- Cartology
- Geodynamic
- Geosurveying
- Photogrammetry
- Graphics
- Physical geo-surveying
- Engineering surveying
- Mapping apparatus

Paper submitted should be sent two printed copies. The main text should be preceded by the abstract of no more than 300 words, followed by key words. Full references should be listed in the order of the citations in the text under the heading “References”, guided by standard publication format. The name of the fund and project series number for articles of funded projects should also be given.

Contributions submitted for publication will be peer reviewed, i.e., the entire articles will be examined by qualified experts independent of the authors before publication.

The authors assume sole responsibility for their dissertations. Canned theses are not allowed. Once the paper is submitted, the editors are authorized to make necessary literate processing. Authors who receive no notice in three months after submission may contact us for inquiry.