Robust Image Denoising Using Kernel-Induced Measures

Keren Tan Songcan Chen Daoqiang Zhang
Nanjing University of Aeronautics & Astronautics,
Nanjing 210016, People's Republic of China.
National Lab of Pattern Recognition, Institute of Automation,
Beijing 100080, People's Republic of China.
tankeren@msn.com s.chen@nuaa.edu.cn daoqz@mail.com

Abstract

In this paper, we propose a class of novel nonlinear robust filters for image denoising by incorporating kernel -induced measures into classical linear mean filter. Particularly, we place more focus on Gaussian kernel based filter (GK) due to its simplicity. The GK filter not only generalizes and makes the original linear mean filter highly resistant to outliers but also outperforms a typical and powerful Mean-LogCauchy filter recently developed by Hamza et al in the mixed noise removal in certain specific conditions in the normalized mean square error (NMSE) sense. Also the experimental results illustrate that the kernel-based nonlinear filters are promising.

1. Formulation of Problem

Image denoising is an important research field of image processing. Its key idea is to build a proper model for underlying random noise distributions in a noisy image and then to use distribution-oriented filters to remove the noise from the image of interest. Among various noise models, the additive noise model is by far the most popular. It can be represented as:

$$X_i = S_i + V_i, \quad i \in Z^n, \tag{1}$$

where $\{S_i\}$ is a discrete n-dimensional deterministic sequence corrupted by the zero-mean noise sequence $\{V_i\}$ and $\{X_i\}$ is the observed sequence. The objective is to estimate the sequence S_i based on $Y_i=F(X_i)$, where F is a filtering operator. We now assume that noise probability distribution P is a scaled version of a member of P_ε such that $P_\varepsilon=(1-\varepsilon)G+\varepsilon S$, where G is Gaussian $N\left(0,\sigma_G^2\right)$, and S is $S\alpha S$ (Symmetric α-Stable) with location θ and dispersion γ_S, and the characteristic function $\varphi(t)$ of P_ε is defined by [1,2]

$$\varphi(t) = \exp(\, j\varepsilon\theta t - (1-\varepsilon)\tfrac{\sigma_G^2}{2}t^2 - \varepsilon^\alpha \gamma_S |t|^\alpha\,), \varepsilon \in [0,1] \tag{2}$$

where the parameter α controls how impulsive the distribution is.

Due to simple computational structure and the efficiency in removing the additive Gaussian noise, linear filters such as the standard mean filter are popular in image denoising. But besides removing noise, these filters also tend to damage image details and perform poorly in the presence of impulsive noise. Nonlinear filters such as the standard median filter are robust, not sensitive to outliers, and preserve image details better in the same case. In practice, the noise is usually not simply pure Gaussian or impulsive noise but a mixed noise with the distribution similar to P, the individual mean or median filter is now not an ideal denoising tool anymore, which has motivated researchers to seek more appropriate filters by using their (or their variant) convex combination to get a performance balance between the mean and median filters in the mixed noise environment. Among them (such as the mean-median filter [2], the mean-relaxed median filter [2], Wilcoxon and Hodges-Lehmann filters [3]), the mean-LogCauchy (MLC) filter recently developed [2], a convex combination of the mean and the LogCauchy filters, has experimentally been proven capable to achieve best performance in removing the mixed noise. Its definition is as follows:

Let W be a sliding window of size $2N+1$, $W_i = \{X_{i+r} : r \in W\}$ be the window data sequence centered at location i. The MLC output is given by

$$Y_i = MLC_\gamma(W_i) = (1-\lambda)\overline{W}_i + \lambda LC_\gamma(W_i) \tag{3}$$

where $\lambda \in [0,1]$, γ is the dispersion of a Cauchy distribution, $\overline{W}_i = \frac{1}{2N+1}\sum_{r=-N}^{N} X_{i+r}$ is the output of the mean filter and given by $Y_i = \overline{W}_i = \arg\min_\theta \sum_{r\in W}(X_{i+r}-\theta)^2$, and the output of the LogCauchy (LC) filter is defined as a solution of the following minimum or equivalently a maximum likelihood estimation problem

$$LC_\gamma(W_i)=\arg\min_\theta J(\theta)\overset{def}{=}\sum_{r\in W}\log\left(\gamma^2+(X_{i+r}-\theta)^2\right) \quad (4)$$

Below for convenience of explanation, we also give the output definition of the median filter $Y_i = \arg\min_\theta \sum_{r\in W}|X_{i+r}-\theta|$. Its equivalent iterative expression can be denoted as $Y_i=\sum_{r\in W}\frac{\frac{X_{i+r}}{|X_{i+r}-Y_i|}}{\sum_{r\in W}\frac{1}{|X_{i+r}-Y_i|}}$.

From the definitions of the above filters, it is not difficult to find that the outputs of the mean, median and *LC* filters result from optimizing the corresponding objective functions associated with the Euclidian, Laplacian and Cauchy measures. In terms of Huber robust statistics [9], it is shown that the latter two measures are robust and can produce nonlinear optimal estimates which can iteratively be solved and at the same time, the former one is not robust but can produces a direct and simple linear analytic estimate. Based on the novel viewpoint of optimizing the measures, we are also now in a position to first derive a class of new robust measures induced by kernels and afterwards present a class of new robust denoising filters based on the new measures.

2. Proposed Filters

2.1 New Measures based on the kernels

Using the Mercer kernels [4], many traditional linear methods have recently been generalized to powerful corresponding nonlinear forms, including principle component analysis [5], *k*-means clustering [6-8] and many applications have successfully demonstrated the power of such kernel methods. Our aim here is also to utilize the kernels to generalize the standard linear mean filters (but not just limited to it) to their corresponding nonlinear versions, more importantly, making it more robust to outliers. To this end, we first will induce a class of new robust (distance) measures for the original space resorting to the kernels and in the sequel develop a class of robust nonlinear iterative filters like the median and *LC* nonlinear filters. To our knowledge, we have not found similar formulation from a viewpoint of the kernel methods. Below is a simple description for the kernel methods aiming at inducing the measures based on the kernel functions.

Let $\Phi:\mathbf{x}\in X\subseteq R^n \mapsto \Phi(\mathbf{x})\in F\subseteq R^H (n\ll H)$ be a nonlinear transformation into a higher (possibly infinite)-dimensional feature space *F*. In order to explain how to use the kernel methods, let us recall a simple example [11]. Assume $\mathbf{x}=\left[x_1,x_2\right]^T$ and $\Phi(\mathbf{x})=\left[x_1^2,\sqrt{2}x_1x_2,x_2^2\right]^T$, where x_i is the *i*th component of vector **x** and ^{T} denotes a transpose of matrix or vector. Then the inner product between $\Phi(\mathbf{x})$ and $\Phi(\mathbf{y})$ in the feature space *F* are: $\Phi(\mathbf{x})^T\Phi(\mathbf{y})=[(x_1)^2,\sqrt{2}\,x_1\,x_2,(x_2)^2][(y_1)^2,\sqrt{2}\,y_1\,y_2,(y_2)^2]^T=(\mathbf{x}^T\mathbf{y})^2=K(\mathbf{x},\mathbf{y})$, hence actually even not knowing emplicitly the mapping $\Phi(\mathbf{x})$, we can also employ some kernel function to directly compute the inner product in *F* [4,5,11,12] defined as follows:

$$K(\mathbf{x},\mathbf{y})\equiv\Phi(\mathbf{x})^T\Phi(\mathbf{y}) \quad (5)$$

The essence behind the method is to realize a more possible linearization for an original complex nonlinear problem in the mapped higher dimensional feature space through a (implicitly) mapping [10, 12] and hence to make more likely the original problem both simple and easy-solving in the latter space.

There are several typical commonly-used kernel functions such as the radial basis function (RBF) kernel:

$$K(\mathbf{x},\mathbf{y})\equiv\Phi(\mathbf{x})^T\Phi(\mathbf{y})=\exp\Big(-\frac{(\sum_{i=1}^n|x_i-y_i|^a)^b}{\sigma^2}\Big)$$

with variance parameter σ and $a\ge 0;\ 1\le b\le 2$ and the polynomial kernel (PK) $K(\mathbf{x},\mathbf{y})\equiv\Phi(\mathbf{x})^T\Phi(\mathbf{y})=(\mathbf{x}^T\mathbf{y}+1)^d$ with the order parameter *d*. For all RBF kernels, $K(\mathbf{x},\mathbf{x})\equiv 1,\forall x$ and when *a*=*b*=1 and *a*=2 and *b*= 1, the RBFs become the exponential (Laplacian) and Gaussian RBFs (GK) respectively. In this paper, we will restrict our kernel to GK for description simplicity. From the above discussion, we define the (Euclidian) distance between **x** any **y** in the feature space as:

$$\begin{aligned}d(\mathbf{x},\mathbf{y})^2&=\|\Phi(\mathbf{x})-\Phi(\mathbf{y})\|^2\\&=\Phi(\mathbf{x})^T\Phi(\mathbf{x})-2\Phi(\mathbf{x})^T\Phi(\mathbf{y})+\Phi(\mathbf{y})^T\Phi(\mathbf{y})\\&=K(\mathbf{x},\mathbf{x})-2K(\mathbf{x},\mathbf{y})+K(\mathbf{y},\mathbf{y})\\&=2-2K(\mathbf{x},\mathbf{y})\end{aligned} \quad (6)$$

The above distance $d(\mathbf{x},\mathbf{y})$ in the feature space corresponds exactly to a class of new non-Euclidian distances in the original space with varying kernels. It has proved in [6, 7] that the measures based on the RBFs including GK are all robust but the measure induced by the PK is not according to Huber M-estimator theory [9]. Furthermore, we can also get an by-product that the RBF kernel-based measures will reduce to general L_p-norm measures so that the L_p-norm-based filters are special cases of corresponding RBF kernels. A brief proof is formulated as follows:

Taking the parameter σ of the RBF kernels as a sufficient large positive constant so that $(\sum_{i=1}^{n}|x_i-y_i|^a)^b \ll \sigma^2$, and from $1-\exp(-x)\le x$ for all non-negative real x, we can obtain

$K(\mathbf{x},\mathbf{y}) \approx 1-[\frac{-(\sum_{i=1}^{n}|x_i-y_i|^a)^b}{\sigma^2}]$ and thus $d(\mathbf{x},\mathbf{y})^2 = 2-2K(\mathbf{x},\mathbf{y}) \approx \frac{2(\sum_{i=1}^{n}|x_i-y_i|^a)^b}{\sigma^2}$, leading to a set of the commonly-used L_p-norm measures when $a=p$ and $b=1/p$. In summary, different kernels can induce different measures with different properties such as robustness and thus can induce new different filters.

2.2 Proposed Filters

Analogously to the derivation of the *LC* and median filters, the output of new filter (called GK filter) with parameter σ in terms of Eq.(6) after some algebra is given by

$$Y_i = \arg\min_{\theta} \sum_{r\in W} d(X_{i+r},\theta)^2$$

$$= \frac{\sum_{r\in W} \exp(-(X_{i+r}-Y_i)^2/\sigma^2)X_{i+r}}{\sum_{r\in W} \exp(-(X_{i+r}-Y_i)^2/\sigma^2)}$$

$$= \frac{\sum_{r\in W} K(X_{i+r},Y_i)X_{i+r}}{\sum_{r\in W} K(X_{i+r},Y_i)} \overset{def}{=} GK(W_i) \qquad (7)$$

This is also a nonlinear iterative solution to Y_i and expressed into a data weighted mean filter like the *LC* filter and this solution is robust in terms of Huber robust statistics [9]. For robustness from Eq. (7), we can give an intuitive explanation. By Eq. (7), the data point X_{i+r} is endowed with an additional weight $K(X_{i+r},Y_i)$, which measures the similarity between X_{i+r} and Y_i. When X_{i+r} is an outlier, i.e., X_{i+r} is far from the other data points, $K(X_{i+r},Y_i)$ will be very small, thus the weighted sum of data points shall be suppressed and hence more robust to outliers. Finally, an iterative solving for Y_i of Eq. (7) starts with an initial value $\overline{W}_i$ (here) and terminates its stable point, i.e., when the absolute of the Y_i

2.3 Mean-GK Filters

Their definition are straightforward like in [2] and we expect the mixture of both can produce better performance than the mean-LogCauchy filters due to the kernel mapping from the low dimensional space to high dimensional feature space. The output of the mean-GK filter (MGK) is

$$Y_i = MGK(W_i) = (1-\lambda)\overline{W}_i + \lambda GK(W_i) \quad (10)$$

where λ and $\overline{W}_i$ are defined as before.

3. Experiment Description

The performances of the MGK filter are evaluated and compared with those of the standard median (SM) and the MLC filter in a mixed noise environments. The original images (256×256 Lena and 256×256 Camera Man) corrupted simultaneously by Gaussian white noise $N(0,100)$ and unit dispersion, zero centered symmetric α-stable (*S*α*S*) noise are used. Restoration performances are quantitatively measured and evaluated by the Normalized Mean Squared Error (*NMSE*) which is defined as

$$NMSE = \sum_{i=1}^{M}\sum_{j=1}^{N}[I(i,j)-R(i,j)]^2 \Big/ \sum_{i=1}^{M}\sum_{j=1}^{N}[I(i,j)]^2 \quad (11)$$

where M and N are the number of image rows and columns respectively, $I(i, j)$ and $R(i, j)$ are the pixel at location (i, j) in the original and reconstructed images respectively.

Filter type (3×3)		SM	MLC_{10}	MGK_{50}
α = 0.5	Lena	0.0102	0.0125	0.0120
	Camera	0.0109	0.0128	0.0125
α = 1.2	Lena	0.0082	0.0077	0.0074
	Camera	0.0093	0.0092	0.0091
α = 1.7	Lena	0.0081	0.0076	0.0073
	Camera	0.0092	0.0091	0.0090

Table 1. NMSEs for mixed Gaussian and α-stable noise

Table 1 compares the performance of these filters using *NMSE* criterion. Experiments in Table 1 are carried out with different α values while the fraction of contamination ε in the noise probability distribution *P* and the parameter λ in MLC and MGK are both set equal to 2/(2+π) [2]. The MLC's parameter γ=10, and the MGK's parameter σ=50 in all experiments in Table 1. From Table 1, it can be seen that the performances of different filters varies with the different α values. When α=0.5, the impulsive behavior and heavy tails of noise are relatively strong. The SM filter gets the best performance. This fact accords with common knowledge about the SM filter's powerful ability to suppress impulsive noises. Although the MGK filter does not get the first place in this case, its performance is still better the MLC filter. As α value increases, the impulsive behavior and heavy tails of noise become gradually weak. The SM filter does not work effectively in these environments. The MGK filter takes the place of the SM filter and achieves the first place in the performance contest. The MLC filter's performance is also very impressive, but it still can not catch up with the MGK filter.

The performances of MGK and MLC filters change with the parameter ε in the noise probability distribution P as showed in Fig 1, where $\lambda = 2/(2+\pi)$ in both filters, and the MLC's parameter γ, the MGK's parameter σ are optimized as much as possible. The definition of *NMSE* difference is given as follows:

NMSE Difference = NMSE(MGK) – NMSE(MLC) (12)

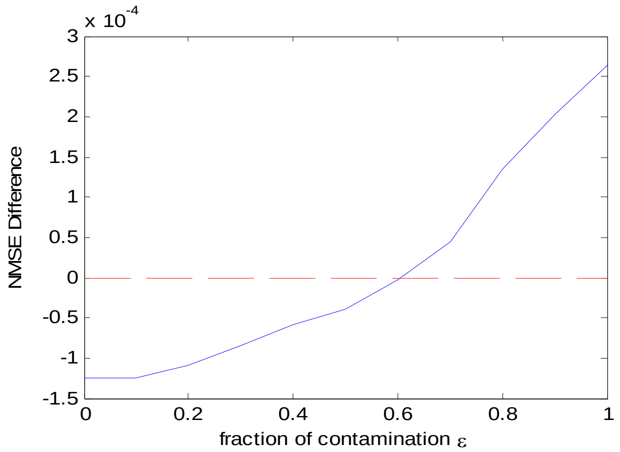

Figure 1. *NMSE* difference between MGK and MLC filters

From this figure, when ε lies in the interval [0, 0.60], the MGK filter outperforms the MLC filter in *NMSE* sense and when ε lies in the interval (0.60, 1], the case is reversed. This phenomenon just proves the important idea in signal processing: every filter has its strengths and limitations and applicable environments, an effective filter should be chosen according to specific environments and requirements — no omnipotent filter exists.

4. Conclusions

We have proposed a class of new robust nonlinear filters based on the Mercer kernels and given a detailed description for the generalized mean filter and comparisons with the Mean- LogCauchy and standard median filters, the results indicate that our proposed approach not only is effective but also outperforms them in NMSE sense in removing the mixed noise in certain specific conditions, which provides a supplement for the weaknesses of the MLC and median filters. What is more, as a newly-developed trick, the kernel method can be used to generalize the existing linear algorithms such as the vector mean even nonlinear vector median filters to their nonlinear counterparts and to make it powerful in the original area.

Acknowledgements

We thank National Science Foundations of China and of Jiangsu under Grant Nos. 60271017 and BK2002092, the "QingLan" Project Foundation of Jiangsu Province and the Returnee Foundation of China Scholarship Council for partial supports respectively.

References

[1] E. E. Kuruoglu, C. Molina, S. J. Gosdill, and W.J. Fitzgerald, *A new analytic representation of the α-stable density function*, in Proc. Amer. Stat. Soc., 1997.

[2] A. Ben Hamza and H. Krim, *Image denoising: A nonlinear robust statistical approach*, IEEE Trans. Signal processing, 49(12), 3045-3053, 2001.

[3] J. Astola and P. Kuosmanen, *Fundamental of Nonlinear Digital Filtering. Boca Raton*, FL: CRC, 1997.

[4] N. Cristianini, J.S. Taylor, *An Introduction to SVMs and Other Kernel-based Learning Methods*, Cambridge University Press, 2000.

[5] B. Schölkopf, A. Smola, and K.R. Müller, *Nonlinear component analysis as a kernel eigenvalue problem,* Neural Computation, 10(5), 1299-1319, 1998.

[6] Daoqiang Zhang and Songcan Chen, *A comment on "Alternative c-means clustering algorithm*", Pattern Recogniton, 37, 173-174, 2004.

[7] Daoqiang Zhang, Songcan Chen, Zhisong Pan, and Keren Tan, *Kernel-based fuzzy clustering incorporating spatial constraints for image segmentation*, IEEE Int. Conf. Mach. Learning & Cybern., Xi'an, China, Nov.2-5, 2003.

[8] J. M. Łęski, *Fuzzy c-varieties/elliptotypes clustering in reproducing kernel Hilbert space*, Fuzzy Sets and Systems, 141(2), 259-280, 2004.

[9] P J. Huber, *Robust statistics*, New York: Wiley, 1981.

[10] T. M. Cover, *Geometrical and statistical properties of systems of linear inequalities in pattern recognition*, IEEE Trans. Electron. Comput., EC-14, 326-334, 1965.

[11] V.N. Vapnik, *Statistical learning theory*. New York: Wiley, 1998.

[12] K.R. Muller, S. Mika, et al, *An Introduction to Kernel-based Learning algorithms*, IEEE Trans. Neural Networks, 12, 181-202, 2001.