Wavelet Denoising Using the Pareto Optimal Threshold

Yifeng Niu and Lincheng Shen

College of Mechatronic Engineering and Automation, National University of Defense Technology, Changsha 410073, China

Summary

The denoising of a natural image corrupted by noise is a classical problem in image processing. In this paper, an efficient algorithm of image denoising based on multi-objective optimization in discrete wavelet transform (DWT) domain is proposed, which can achieve the Pareto optimal wavelet thresholds. First, the multiple objectives for image denoising are presented, then the relation between these objectives and the wavelet thresholds is analysed, finally the algorithm of adaptive multi-objective particle swarm optimization is introduced to optimize the wavelet thresholds. Experiments show that the Pareto optimal threshold-denoising algorithm is more effective than other algorithms, and can attain the Pareto optimal denoised image.

Key words:

Image denoising, multi-objective optimization, particle swarm optimization, discrete wavelet transform, thresholding function.

1. Introduction

Threshold selection is the critical issue in image denoising via wavelet shrinkage. Many powerful approaches have been investigated, and the most well-known thresholds include the universal threshold [1, 2], the adaptive threshold [3], and the Bayesian threshold [4], where Bayesian performs well most of the time. However, few approaches have made the threshold values optimal for the noisy images. In general, the thresholds are given or adaptively change based on the image contents, so it is fairly difficult to obtain the optimal denoised image. In fact, image denoising can be regarded as an optimization problem, and the optimal result can be acquired by searching the optimal thresholds. Therefore, the proper optimization objectives and search strategy are highly important. Shi [5] primarily explored the problem, and introduced the Nelder-Mead simplex method to search the optimal threshold of image denoising, the objective is PSNR (peak signal-to-noise ratio). However the method can't meet the real demands.

Actually the evaluation criteria of image quality can be used as the optimization objectives. However, these criteria are various, and the different criteria may be are compatible or incompatible with one another. The conventional solution is to change the multi-objective problem into a single objective problem using weighted linear method. However, the relation of the criteria is often nonlinear, and this method needs to know the weights of different criteria in advance. So it is highly necessary to introduce evolutionary multi-objective optimization methods based on Pareto theory [6] to search the optimal thresholds in order to realize the optimal image denoising. In this paper, adaptive multi-objective particle swarm optimization (AMOPSO) [7] is introduced and applied to optimize the wavelet thresholds of image denoising. Experiments show that the approach to image denoising based on AMOPSO is more successful.

2. Wavelet Thresholding Function

The wavelet thresholding procedure removes noise by thresholding only the wavelet coefficients of the detail subbands, while keeping the low resolution coefficients unaltered. There are four thresholds frequently used, i.e. a hard threshold, a soft threshold, a semi-soft threshold, and a nonlinear soft-like threshold.

The hard-thresholding function keeps the input if it is larger than the threshold; otherwise, it is set to zero [1]. It is described as

$$\eta_1(w) = wI(|w| > T) \quad (1)$$

where w is a wavelet coefficient, T is the threshold, and $I(x)$ is a function the result is one when x is true and zero vice versa.

The soft-thresholding function (also called the shrinkage function) takes the argument and shrinks it toward zero by the threshold [2]. It is described as

$$\eta_2(w) = (w - \mathrm{sgn}(w)T)I(|w| > T) \quad (2)$$

where sgn(x) is the sign of x. The soft-thresholding rule is chosen over hard-thresholding, for the soft-thresholding method yields more visually pleasant images over hard-thresholding.

The semi-soft thresholding function is a more general shrinkage function [8] and described as

$$\eta_3(w) = \mathrm{sgn}(w)\frac{T_2(|w| - T_1)}{T_2 - T_1}I(T_1 < |w| < T_2) + wI(|w| > T_2) \quad (3)$$

where $T_2 > T_1 > 0$. By choosing appropriate thresholds, the semi-soft shrinkage offers advantages over both hard thrinkage (uniformly small risk and less sensitivity to small perturbations in the data) and soft shrinkage (small bias and overall risk).

Since the soft-thresholding functions are continuous with discontinuous derivative. However, the continuous

Manuscript received January 1, 2007
Manuscript revised January 25, 2007

derivative or higher derivatives are often desired for optimization problems. The nonlinear soft-like thresholding function is constructed with continuous derivatives [9]. It is described as

$$\eta_4(w)=\begin{cases} w+T-T/(2k+1), & w<-T \\ w^{2k+1}/((2k+1)T^{2k}), & |w|\le T \\ w-T+T/(2k+1), & w>T \end{cases} \quad (4)$$

where k is a positive integer. When k is near to $+\infty$, the limit is just the commonly used soft-thresholding function. The nonlinear soft-like thresholding functions have better numerical properties.

3. Criterion Establishment

In our method of image denoising, the establishment of an evaluation criterion system is the basis of the optimization that determines the quality of the denoised image. In fact, the evaluation criteria of image denoising can be divided into two categories. One category reflects the image features, such as entropy and variance; the second reflects the relation of the fused image to the reference image, such as peak signal to noise ratio (PSNR) and structural similarity.

Entropy is an criterion to evaluate the information quantity contained in an image. A noisy image always has high entropy. If the value of entropy becomes lower after denoising, it indicates that the performance of image denoising is improved. Entropy is defined as

$$H=-\sum_i p_i \log p_i \quad (5)$$

where $p=\{p_0, p_1, \ldots, p_{L-1}\}$ is the probability distribution of each grey level, L is the total of levels. The maximum value of entropy in a grey-scale image is

$$H(F)\le \log L \quad (6)$$

If $p_1=p_2=\ldots=p_n=1/L$, the equality will hold.

Standard deviation (SD) reflects the deviation of image grey contrast to the mean. The high the value of standard deviation is, the more dispersive the distribution of grey level is, and the stronger the noise is. The standard deviation of σ is defined as

$$\sigma^2=\frac{1}{MN-1}\sum_{i=1}^{M}\sum_{j=1}^{N}[F(i,j)-\mu]^2 \quad (7)$$

where M and N are the numbers of the row and the column of the image respectively, μ is the mean value of the image grey.

$$\mu=\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}F(i,j) \quad (8)$$

The higher the value of PSNR (Peak Signal to Noise Ratio) is, and the lower the value of RMSE is, the better the fused image is. PSNR is defined as

$$PSNR=10\lg\frac{255^2}{RMSE^2} \quad (9)$$

where RMSE (root mean squared error) is defined as

$$RMSE^2=\frac{1}{MN}\sum_i\sum_j[R(i,j)-F(i,j)]^2 \quad (10)$$

Simulated the human vision system, structural similarity (SSIM) [10] is designed by modelling any image distortion as a combination of three factors: structure distortion, luminance distortion, and contrast distortion. It is defined as

$$SSIM=\frac{\sigma_{FR}+C_1}{\sigma_F\sigma_R+C_1}\cdot\frac{2\mu_F\mu_R+C_2}{\mu_F^2+\mu_R^2+C_2}\cdot\frac{2\sigma_F\sigma_R+C_3}{\sigma_F^2+\sigma_R^2+C_3} \quad (11)$$

where μ_F and μ_R are the mean intensity of the fused image F and the reference image R respectively, σ_F and σ_R are the standard deviation of F and R, σ_{FR} is the covariance, C_1, C_2, and C_3 are positive constant to avoid instability when denominators are very close to zero. In (11), the first component is the correlation coefficient of F and R. The second component measures how close the mean grey levels of F and R is, while the third measures the similarity between the contrasts of F and R. The higher the value of SSIM is, the more similar to R the F is. The dynamic range of SSIM is [-1, 1]. If two images are identical, the similarity is maximal and equals 1; while one is the negative of the other, SSIM equals -1.

4. AMOPSO Algorithm

Kennedy and Eberhart brought forward particle swarm optimization (PSO) inspired by the choreography of a bird flock in 1995 [11]. PSO possesses better optimization capacities and has shown a high convergence speed in both single objective optimization and multi-objective optimization [12]. In order to improve the performances of the algorithm, we presented a proposal, called "adaptive multi-objective particle swarm optimization algorithm" (AMOPSO) [7], in which a new crowding operator based on Manhattan distance is used to improve the distribution of nondominated solutions along the Pareto front and maintain the population diversity; an adaptive inertia weight and the adaptive mutation operator are introduced to improve the search capacities and avoid the earlier convergence; the uniform design [13] is used to obtain the optimal combination of the algorithm parameters. The algorithm has been successfully applied in multi-objective optimization problems.

The algorithm of AMOPSO is the following:

Step 1. Initialize the position of each particle: *pop*[*i*]=arbitrary, where $i=1,\ldots,N_P$, N_P is the particle number; initialize the velocity of each particle: *vel*[*i*]=0; initialize the record of each particle: *pbests*[*i*]=*pop*[*i*]; evaluate each of the particles in the *POP*: *fun*[*i, j*], where $j=1,\ldots,N_F$, and N_F is the objective number; and store the positions that represent nondominated particles in the repository of the *REP* according to the Pareto optimality.

Step 2. Update the velocity of each particle using the following expression.

$$vel[i] = W \cdot vel[i] + c_1 \cdot rand_1 \cdot (pbests[i] - pop[i]) + c_2 \cdot rand_2 \cdot (rep[h] - pop[i]) \quad (12)$$

where W is the adaptive inertia weight; c_1 and c_2 are the learning factors, $rand_1$ and $rand_2$ are random values in the range [0, 1]; $pbests[i]$ is the best position that particle i has had; h is the index of the maximum crowding distance in the repository that implies the particle is located in the sparse region, as aims to maintain the population diversity; $pop[i]$ is the current position of particle i.

Step 3. Update the new positions of the particles adding the velocity produced from the previous step.

$$pop[i] = pop[i] + vel[i] \quad (13)$$

Step 4. Maintain the particles within the search space in case they go beyond their boundaries (avoid generating solutions that do not lie in valid search space).

Step 5. Adaptively mutate each of the particles in the *POP* at a probability of P_m.

Step 6. Evaluate each of the particles in the *POP*.

Step 7. Update the contents in the *REP*, and insert all the current nondominated positions into the repository.

Step 8. Update the record of each particle. When the current position of the particle is better than the position contained in its memory, the latter is updated.

$$pbests[i] = pop[i] \quad (14)$$

Step 9. If the maximum cycle number is reached, stop the process and output the Pareto solutions; else go to Step 2.

5. The Pareto Optimal Wavelet Denoising

Our Pareto optimal wavelet denoising method can be summarized as follows:

1) Select the proper thresholding function for image denoising; and construct the objective functions, including entropy, SD, PSNR, and SSIM;
2) Perform a 2D discrete wavelet transform for the noisy image to get the noisy wavelet coefficients in *LL*, *LH*, *HL*, and *HH* components;
3) Optimize the thresholds of the thresholding function using AMOPSO; when the maximum cycle is reached, output the Pareto solutions in the repository; according to the preference, select the Pareto optimal thresholds;
4) Threshold the coefficients in *LH*, *HL*, and *HH* components with the Pareto optimal thresholds, and get the modified new *HL'*, *LH'*, and *HH'* components;
5) Reconstruct the denoised image with *LL*, *LH'*, *HL'*, and *HH'* components.

6. Results

We do some experiments to test the proposed thresholding algorithm. First, we analyze the relations between the standard deviation of the noise and evaluation criteria, so as to verify the criteria sound. The test image set includes Lena, Baboon, Couple, Elaine, and Man. Suppose that the standard deviation is no greater than 40. As shown in Fig. 1, the noise-free images have the same variety curves with the strength of the Gaussian white noise, and these criteria can effectively evaluate the quality of a denoised image. So we select "Lena" to do the further experiments.

Second, we analyze the relations between the wavelet threshold and evaluation criteria. In order to get general results, the soft-thresholding function is selected; the decomposition level is set to one. As shown in Fig.2, each criterion can reach its extremum at some threshold, while these thresholds are usually unequal. Thus, we must introduce the Pareto theory to optimize the threshold and balance the advantages of the conflicting criteria.

Third, we do the experiments on the Pareto optimal wavelet threshold for image denoising with the reference image, where the number of objectives is four, and the standard deviation of the noise is known. The image Lena of size 512×512 is added Gaussian noise, the decomposition level is set to one. The principal objective is SSIM, then PSNR, SD, and Entropy. By the uniform design, the parameters of AMOPSO are as follow: the particle number of N_P is 100; the maximum cycle number of G_{max} is 100; the allowed maximum capacity of *MEM* is 200; the mutation probability of P_m is 0.06. The results are shown in Table 1. It can be seen that the Pareto optimal nonlinear thresholding is the best denoising algorithm.

Finally, we do the experiments on the Pareto optimal wavelet threshold for image denoising without the

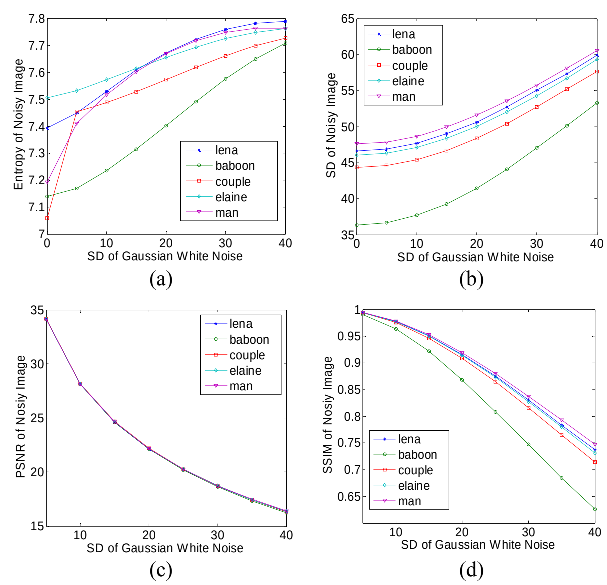

Fig. 1 The relations between SD of the noise and evaluation criteria

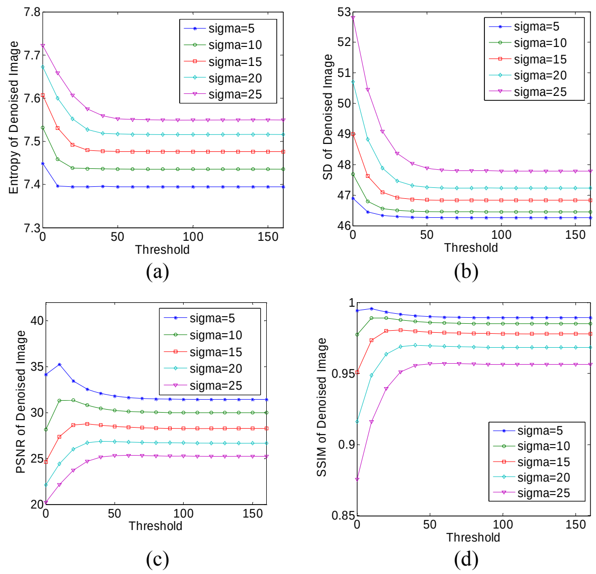

Fig. 2 The relation between the wavelet threshold and evaluation criteria

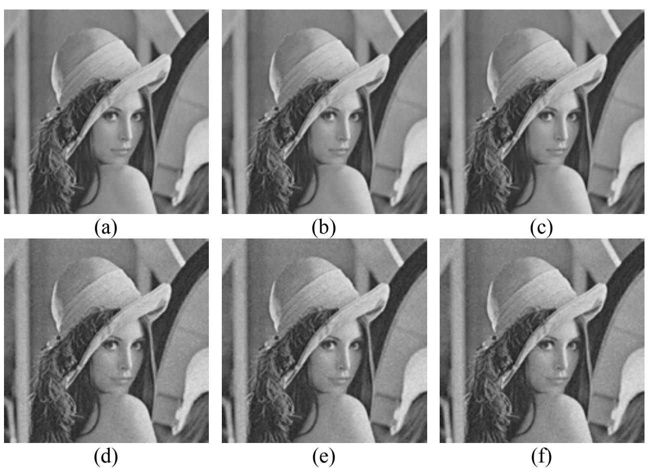

Fig. 3 The denoised images: (a) Bayesian denoised result (σ=10); (b) Our result with the reference image (σ=10); (c) Our result without the reference image (σ=10); (d) Bayesian denoised result (σ=20); (e) Our result with the reference image (σ=20); (f) Our result without the reference image (σ=20)

reference image, where the number of objectives is two, the standard deviation of the noise is unknown. Thus, SD of the noise from the wavelet coefficients should be estimated. A common experimental estimator [1] is as follows

$$\hat{\sigma}_n = Mad / 0.6745 \qquad (16)$$

where *Mad* is the median of the absolute values of the wavelet coefficients.

The principal objective is SD, then entropy. Finally, the denoised image is compared to the standard image, to get the values of PSNR and SSIM. The results are shown in

Table 2. The same conclusion can be drawn, i.e. the nonlinear threshold can get the Pareto optimal objectives.

Table 1: The criteria for different denoising algorithms with the reference noisy image

Schemes		Sigma=10			
		Entropy	SD	PSNR	**SSIM**
Universal		7.3875	45.3684	31.4152	0.9890
Adaptive		7.3875	45.3678	31.3877	0.9889
Bayesian		7.3891	45.3785	31.6701	0.9896
Pareto Optimal Threshold	Hard	7.3892	45.4469	31.6568	0.9896
	Soft	7.4020	45.4290	31.9041	0.9901
	Semi-Soft	7.4021	45.4374	31.9048	0.9901
	Nonlinear	7.4072	45.4354	31.9543	0.9902
Schemes		Sigma=15			
Universal		7.4276	45.7320	29.2412	0.9819
Adaptive		7.4276	45.7319	29.2395	0.9819
Bayesian		7.4282	45.7592	29.2552	0.9820
Pareto Optimal Threshold	Hard	7.4278	45.7463	29.2695	0.9820
	Soft	7.4351	45.7618	29.3534	0.9824
	Semi-Soft	7.4329	45.7605	29.3435	0.9824
	Nonlinear	7.4422	45.7577	29.3932	0.9825
Schemes		Sigma=20			
Universal		7.4651	46.0894	27.3061	0.9719
Adaptive		7.4651	46.0893	27.3060	0.9719
Bayesian		7.4652	46.0910	27.3076	0.9720
Pareto Optimal Threshold	Hard	7.4652	46.0910	27.3076	0.9720
	Soft	7.4687	46.1033	27.3243	0.9721
	Semi-Soft	7.4664	46.0991	27.3178	0.9720
	Nonlinear	7.4776	46.1019	27.3490	0.9722

Table 2: the criteria for different denoising algorithms without the reference noisy image

Schemes		Sigma=10		Evaluation	
		Entropy	**SD**	PSNR	SSIM
Universal		7.3875	45.3680	31.3999	0.9889
Adaptive		7.3875	45.3678	31.3861	0.9889
Bayesian		7.3892	45.3740	31.6176	0.9895
Pareto Optimal Threshold	Hard	7.3875	45.3678	31.3847	0.9889
	Soft	7.3875	45.3678	31.3847	0.9889
	Semi-Soft	7.3875	45.3678	31.3847	0.9889
	Nonlinear	7.3911	45.3624	31.6041	0.9894
Schemes		Sigma=15		Evaluation	
Universal		7.4276	45.7320	29.2406	0.9819
Adaptive		7.4276	45.7319	29.2395	0.9819
Bayesian		7.4276	45.7319	29.2395	0.9819
Pareto Optimal Threshold	Hard	7.4276	45.7319	29.2395	0.9819
	Soft	7.4276	45.7319	29.2397	0.9819
	Semi-Soft	7.4276	45.7319	29.2397	0.9819
	Nonlinear	7.4285	45.7286	29.2630	0.9820
Schemes		Sigma=20		Evaluation	
Universal		7.4651	46.0894	27.3060	0.9719
Adaptive		7.4651	46.0893	27.3060	0.9719
Bayesian		7.4651	46.0894	27.3060	0.9719
Pareto Optimal Threshold	Hard	7.4651	46.0894	27.3060	0.9719
	Soft	7.4652	46.0894	27.3070	0.9720
	Semi-Soft	7.4651	46.0894	27.3062	0.9719
	Nonlinear	7.4672	46.0886	27.3188	0.9720

The denoised images with different algorithms are shown in Fig. 3. Bayesian threshold is the best wavelet threshold, so the results from Bayesian thresholding are compared with our denoising results. It can be seen that the Pareto wavelet threshold for image denoising can achieve the better results, whether the reference image exists or not. Especially when the noise variance is unknown in advance, the Pareto optimal wavelet thresholding is reasonably superior.

7. Conclusion

We have demonstrated that evolutionary multi-objective optimization can be used for efficient image thresholding denoising. The performance can be improved further by exploiting the curvelet transform and other superior thresholding functions.

References

[1] Donoho D.L. and Johnstone I.M., "Ideal spatial adaptation via wavelet shrinkage", Biometrika, 1994, 81, (3), pp. 425-455

[2] Donoho D.L., "De-noising by soft-thresholding", IEEE Trans. Inform. Theory, 1995, 41, (3), pp. 613-627

[3] Chen Y. and Han C., "Adaptive wavelet threshold for image denoising", Electron. Lett., 2005, 41, (10), pp. 586-587

[4] Chang S.G., Yu B., and Vetterli M., "Adaptive wavelet thresholding for image denoising and compression", IEEE Trans. Image Process., 2000, 9, (9), pp. 1532-1546

[5] Shi G.M. and Li F.D., "Image denoising with optimized subband threshold", Proc. Int. Conf. Computational Intelligence and Multimedia Applications (ICCIMA), Xi'an, 2003, pp. 160-164

[6] Coello Coello C.A., "Recent trends in evolutionary multiobjective optimization", in A. Abraham, et al. (Eds.): Evolutionary Multiobjective Optimization: Theoretical Advances and Applications, Springer-Verlag, London, 2005, pp. 7-32

[7] Niu Y.F. and Shen L.C., "Multi-resolution image fusion using AMOPSO-II", in: D.S. Huang, et al. (Eds.): Intelligent Computing in Signal Processing and Pattern Recognition, LNCIS 345, Springer-Verlag, Berlin Heidelberg, 2006, pp. 345-354

[8] Gao H.Y. and Bruce A.G., "WaveShrink with firm shrinkage", Statist. Sinica, 1997, 7, (4), pp. 855-874.

[9] Zhang X.P. and Desai M., "Adaptive denoising based on SURE risk", IEEE Signal Process. Lett., 1998, 5, (10), pp. 265-267

[10] Wang Z., Bovik A.C., Sheikh H.R., et al, "Image quality assessment: from error visibility to structural similarity", IEEE Trans. Image Process., 2004, 13, (4), pp. 600-612

[11] Kennedy J. and Eberhart R.C., "Particle swarm optimization", Proc. Int. Conf. Neural Networks, Perth, 1995, 4, pp. 1942-1948

[12] Reyes-Sierra M. and Coello Coello C.A., "Multi-objective particle swarm optimizers: a survey of the state-of-the-art", Int. J. Comput. Intell. Research, 2006, 2, (3), pp. 287-308

[13] Leung Y.W. and Wang Y.P., "Multiobjective programming using uniform design and genetic algorithm", IEEE Trans. Syst. Man Cybern. Pt. C: Appl. Rev. 2000, 3, pp. 293-304

Yifeng Niu received the B.S. degree from National University of Defense Technology (NUDT), Changsha, China, in 2001. He is currently a Ph.D. candidate of Automation Institute in NUDT. His current research interests include image processing and information fusion

Lincheng Shen received the B.S., M.S., and Ph.D. degree from National University of Defense Technology (NUDT), Changsha, China, in 1986, 1989 and 1994. He is currently a professor of College of Mechatronic Engineering and Automation in NUDT. His current research interests include artificial intelligence, image processing, and information fusion.