

The Heuristic Algorithm of Wavelet Image Denoising Based on Rough Set

Song Lan^{1,2}, Lei Lixia², Li Haisheng² and Hong Wang³

¹State Key Laboratory of Software Engineering, Wuhan University, Wuhan 430072, Hubei, China

²Information and Engineering College, East China JiaoTong University, Nanchang 330013, Jiangxi, China

³Nanchang city sports bureau, Nanchang 330000, Jiangxi, China
sl130com@gmail.com, llx163@163.com, lhs@ecjtu.jx.cn, jack108@163.com

Abstract

In this paper, we propose a novel approach to explore image denoising for patch based image process. The importance measurement model of Rough Entropy and the importance reduction method of wavelet coefficients are given. This paper combines the rough set theory with the denoising method of wavelet threshold, regarding the high-frequency information blocks in the transform domain as similar ones, and adopting importance Reduction Methods to contract the coefficients. The simulation results show that this method is effective.

Keywords: Image denoising; Rough set; Wavelet coefficient; Threshold

1. Introduction

The original image collected in the actual project contains a certain amount of noise. The noise-contained image is of poor quality, so it fails to reflect the information contained in real image. Removing the noise is a critical technology of wavelet image processing: the wavelet transform can decompose the image signal into characteristic coefficients with different resolutions and the latter can then be analyzed and processed in order to remove noise. In the year of 1994, Donoho and other scholars [1, 2] put forward the wavelet shrinkage algorithm for image denoising. It can achieve the goal of image denoising by removing the wavelet coefficient which is less than a certain threshold. The most commonly-used method is to remove or reduce these high-frequency coefficients in the transform domain, and then inverse transform to get the image, so as to achieve the purpose of denoising.

The estimate of the threshold is a key issue of the wavelet threshold algorithm for image denoising. If the threshold is too small, the noise will still exist in denoised image. On the contrary, if the threshold is too big, some important image features will be filtered out. In practical project, Hard Threshold [1], Soft Threshold [2] and other methods are usually used to determine the threshold. With a certain amount of guesswork, these commonly used methods tend to have an unstable effect of denoising.

Buades and other researchers who makes comparative study of many typical de-noising algorithm proposes NLM (Non-ILocal Means) and applies it into the denoising of images and videos [3]. NLM was restricted to the local neighborhood as in a bilateral filter [4]. The basic idea of this method is to remove noise and eliminate the images artifacts which usually occur in traditional filtering algorithm by using the “repeating structure” property of images. NLM algorithm uses a Weighted Euclidean Distance between gray value vectors to measure the similarity between the pixel neighborhoods. NLM algorithm uses the same filtering

parameters on the whole image, therefore the denoising effect depends on the selection of the filtering parameters. When the noise level exceeds a certain value, it is difficult to find a global filtering parameter which can better denoise various parts of the image [5].

Up till now, there are two types of de-noising method based on the block: one is the weighted average denoising of similar blocks among the candidate blocks; the other is the conversion domain denoising. In 2007, Dabov K, Foi A, Katkovnik V and Egiazarian KO proposed a three-dimensional block matching noising method (BM3D). Based on the idea of non-local mean, this method first finds the similar blocks among the to-be-denoised images, then gathers these similar blocks to form a three-dimensional matrix blocks, finally combines filtering. The denoising is realized in the transform domain [6]. Due to the images in the wavelet domain processing was only about a quarter of the time domain, through the noisy image to the wavelet transform domain, faster processing speed will be obtained.

The similar blocks based Image matching is a very time consuming work, By using Rough Set Theory and the concept of attribute significance, put forward a new method to study the similarity between image blocks. According to the concept of importance of attribute, this paper analyzes the approximate and detail components at different scales got by the wavelet transform of the original image and divide the wavelet coefficients into several blocks, then uses the coefficient of horizontal, vertical and diagonal line for similarity judgment. Finally adopts a knowledge-reduction-based algorithm according to its importance in the signal classification in order to achieve adaptive threshold selection and the purpose of image denoising. The results of the simulation experiments prove that this method can achieve better results.

2. Rough Entropy and the Concept of Attribute Significance

Rough Entropy is a mathematical theory put forward by a Poland scientist Z. Pawlak in the early 1980s. It is an important branch of the field of artificial intelligence. It is mainly used to describe the problem of uncertainty which fails to be illustrated by classic collection theory. Rough Entropy can directly proceed from the given data sample, and find out the inner rule which is implicit in the data through the approximate domain (Upper and Lower Approximation) of problems which are identified by undistinguishable relations and equivalence partitioning [7].

The Knowledge Reduction is a core part of Rough Set Theory. We must first determine whether the knowledge is redundant. In the Rough Set Theory, it is very easy to determine whether the knowledge is redundant or not: after a certain knowledge is removed from the knowledge base, we should examine whether the classification ability of this knowledge is weakened or not. If it is weakened, it indicates that the knowledge is useful; otherwise, this knowledge is redundant.

Definition 1: Supposing that $S=(U,A)$ is an information system, U is a finite set of objects, A is a finite set of attributes, $P \subseteq A$, then the Rough entropy of knowledge P is defined as:

$$E(P) = - \sum_{i=1}^n Q_i \log P_i \quad (1)$$

Of which,

$$P_i(x) = 1/|D(S_p(x_i))|, \quad Q_i(x) = |D(S_p(x_i))|/|U|, \quad | \cdot | \text{ is the potential of collection.}$$

$D(S_p(x_i))$ represent all possible values of x_i -similar objects in the decisional attribute.

Definition 2: Supposing that $T=[U,A,V,F]$ is a decisional table, and $P \subseteq A$, then the following proposition holds:

- (1) $E(P)$ takes the maximum value $|U| \cdot \log_2 |U|$, if and only if $U/P = \{U\}$;
- (2) $E(P)$ takes the minimum value of 0, if and only if $U/P = \{\{x_1\}, \{x_2\}, \{x_3\}, \dots, \{x_{|U|}\}\}$.

After analyzing the nature of rough entropy about knowledge, we drew the conclusions as follows: Taking Knowledge P as an example, if we cannot distinguish between any two objects in the domain, then the roughness of Knowledge P ranks the biggest; if we can distinguish between any object in the domain, then Knowledge P achieves the most accurate extent[8].

Definition 3: Supposing $T=[U,C \cup D,V,F]$ is a decisional table, C is a set of conditional attributes, D is a set of decisional attributes, $a \in C, A=C \cup D$. We can define the importance of conditional attribute A to the conditional attribute set C as follows:

$$SIG_{A-\{a\}}(a) = E(A-\{a\}) - E(A) \tag{2}$$

The above definition shows the importance of the attribute $a \in A$ to the set of attributes A is decided by the change of amount of information caused by removing $a \in A$ from A. The larger the value of $SIG_{A-\{a\}}(a)$ is, given the fact that A is already known, the more important the attribute $a \in A$ is for decision-D.

As a measurement of the attribute significance, Rough Entropy deletes irrelevant or unimportant knowledge on the condition that the classified or decision-making ability of knowledge database remain unchanged, so as to achieve the purpose of knowledge reduction. As the Rough entropy of knowledge decreases monotonically with the enhanced ability to distinguish, we can set the adaptive threshold value of the transformed image's wavelet coefficients.

3. Basic Theory on the Denoising of Wavelet Threshold

The mathematical model [9] of the denoised image can be expressed as follows:

$$g(x, y) = h(x, y) + n(x, y) \tag{3}$$

Of which, g is the observed image, h is a true image, n is the noise added in the real image.

After the scale decomposition, Formula (3) can be expressed as follows:

$$w_{j,k} = \hat{w}_{j,k} + n_{j,k} \tag{4}$$

$w_{j,k}$ represents the wavelet coefficients containing the noise, $\hat{w}_{j,k}$ represents the denoised wavelet coefficients.

The principle of the denoising method of wavelet threshold is as follows: when $w_{j,k}$ is less than a certain threshold, it will be removed as the noise; when $w_{j,k}$ is larger than a certain threshold, it will not be removed. For a given wavelet coefficients, the greater the noise is, the

greater the value of the threshold is. Therefore, the threshold value can be calculated based on the statistical characteristics of a set of wavelet coefficients.

The selection of the threshold value has to be able to meet the minimum mean-square error, *i.e.*,

$$\lambda_{MSE} = E \left\| \hat{w}_{j,k} - w_{j,k} \right\|_F \quad (5)$$

λ_{MSE} formula has to reach the minimum in order to make the obtained denoised image as close as possible to the original image.

Documents [10, 11] proved that: in order to make the λ_{MSE} formula (5) reaches the minimum, it must satisfy the following formula:

$$\lambda_{MSE} \square \sigma \sqrt{2 \log(N)} \quad (6)$$

For the standard deviation σ of the noise, the method of estimation is usually adopted. The high-frequency detail sub-band of noise-contained image usually contains little image information, and its energy is mainly provided by the noise. Therefore, Donoho proposed to estimate the variance of the noise in accordance with the first layer of detail coefficients.

$$\sigma = \frac{\text{median}(|w_{1,k}|)}{0.6745} \quad (7)$$

Of which, median is the mean value function, the standard deviation of the noise is got by diving the absolute deviation of mean value by 0.6745.

4. The Heuristic Algorithm of Wavelet Denoising based on Rough Entropy

NLM, as the latest methods, discuss the application of ROUGH SET Theory into the search of global similar blocks in images. However, these methods are greatly time-consuming. One of the main reasons is that the calculation of each candidate block has to be done. If we classify the candidate blocks into different groups in advance and then calculate for each group, we will be possibly able to speed up image processing [12].

Granularity is an important concept in ROUGH. We will apply this method to achieve image denoising. The image space can be partitioned in non-overlapping small granules to derive more information or for constituting class defined by attributes for each object present in the image. Lower Approximation consists of a set of certain class defined by attributes, while the Upper Approximation consists of a set of possible class defined by attributes. The division of Upper- and Lower- Approximation of objected presented in the images can be achieved by setting corresponding upper- and lower- threshold. The block size takes the form of square, and the specific size is determined according to experience.

If a candidate patch is within a lower approximation of an object, then patches which constitute the lower approximation of this object are similar patches. If a patch is in the upper approximation of an object but not in its lower approximation, then the patch is in the boundary region of two distinct objects. Any patch that is said to be similar to this patch must have the same property.

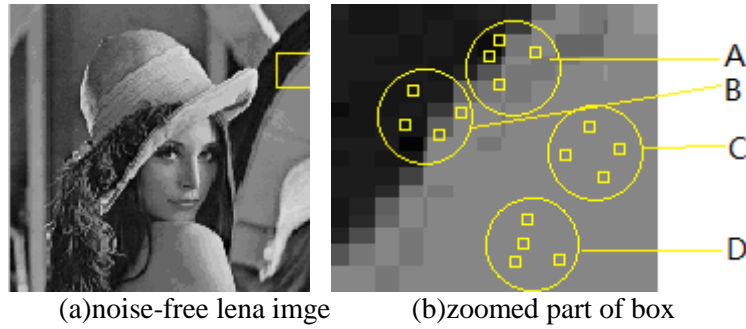


Figure 1. Patch Selection using Rough Set

A noise-free lena image is shown in (a) and zoomed part of box is shown in (b). In (b), groups of patches are shown in circle centered at patches A-D. The neighbor of patches A and B is having the same property and belongs to the two different object as shown. For denoising them one can use patches which are falling on the boundary of two same object. The neighbor of patches C and D is having the same property and belongs to the same object. Thus, For denoising patch C, one can use the patch D or vice versa.

First, we carry out the multi-scale wavelet transform of image $g(x,y)$, analyze the different-characterized wavelet coefficient $W_{j,k}$ (a coefficient in scale 2^j) by taking the noise $n(x,y)$ into consideration, conduct the different degrees of separation between high-frequency component and low-frequency component in the original image by using the wavelet decomposition, then extract the wavelet energy which is the most sensitive to the characteristics of image by using the Knowledge Reduction Theory of Rough Set, finally improve the image effect by using the method of wavelet reconstruction after setting the corresponding threshold processing function and processing the noise. The process is illustrated in Figure 2.

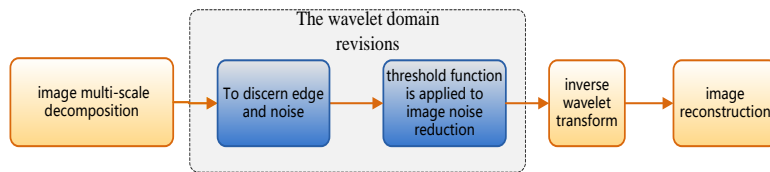


Figure 2. Image Denoising Method based on Wavelet

The wavelet coefficients energy in scale J is defined as $U_j = \|W_j\|^2 = \sum_{j=1}^n |w_{j,k}|^2$, and the total energy of image is defined as $U = \sum_{j=1}^m U_j$.

The new processing function of wavelet soft threshold [13] is as follows:

$$\hat{w}_{j,k} = \begin{cases} w_{j,k} - \lambda \operatorname{sgn}(w_{j,k}) \exp(-(U \log U - E)(|w_{j,k}| - \lambda)), & |w_{j,k}| \geq \lambda \\ 0, & |w_{j,k}| < \lambda \end{cases} \quad (8)$$

Of which

(1) $-(U \log U - E)$ shows the correlation between the entropy and the processing function of wavelet coefficient threshold; $w_{j,k}$ and $\hat{w}_{j,k}$ represents the decomposed wavelet coefficient and the corresponding denoised wavelet coefficient respectively.

(2) Threshold λ_j represents the threshold of wavelet coefficient in layer j [14].

$$\lambda_j = \frac{\sigma \sqrt{2 \log N}}{\log(j+1)} \quad (9)$$

(3) the variance of the estimated value

$$\sigma = \frac{\text{median}(|w_{\min ent}|)}{0.6745} \quad (10)$$

$w_{\min ent}$ is the set of detail wavelet coefficients with the smallest entropy value after the conversion of the first layer.

By using the measurement of the attribute significance in the Rough Set Reduction Theory given in Section 2, we can extract the wavelet coefficient which is the most sensitive to the characteristics of images. In this section, the writer puts forward a heuristic algorithm. Its basic idea is as follows: it regards the approximate decomposed wavelet coefficient as the core attribute, calculates the importance of the inter-cell component of wavelet coefficient to the core attribute, then find out the maximum value and classified it as a member of the core attribute, eventually realized the reduction of attribute set.

Algorithm:

Input: a decisional table $T=\{U,C \cup D,V,F\}$, of which, U is the domain of discourse, C and D are the conditional attribute set and decisional attribute set respectively.

Output: a relative reduction of the decisional table

Step 1: dividing the transformed wavelet coefficient into several small sections. The coefficient of the whole sections is regarded as the set of conditional attribute C. Calculating the entropy $E(a_i)$ of each small section. a_1, a_2, \dots, a_n are the wavelet coefficient energy in different sections.

Step 2: regarding the decomposed low-frequency wavelet coefficient as the core attribute B, $Att=C - B$.

Step 3: if $E(B) \neq kE(C)$, then

begin

- (1) For each attribute $a_j \in Att$, calculate the Rough entropy $E(B \cup \{a_j\})$ of decisional attribute D with respect to the conditional attribute set $B \cup \{a_j\}$
- (2) Choose the attribute with the maximum value of $E(B \cup \{a_j\})$ in Att, $Att=Att - \{a\}, B=B \cup \{a\}$
- (3) If $E(B) = kE(C)$, then move on to Step 4; otherwise, move on to Step 1
- (4) Output a reduction B of decisional table.

end

The algorithm sets the core attribute set of decisional table as a starting point, adds successively the non- core attribute A which makes the minimum value of $E(B \cup \{a_j\})$ to the core attribute set until it satisfies the termination condition $E(B) = kE(C)$, finally get the reduction results.

Given the worst case, the number of attributes is $n, n-1, \dots, 1$ (n is the number of wavelet coefficient), the total number is $n+(n-1)+ \dots +1=n(n+1)/2$. Therefore, in the worst case, this algorithm can help to find a satisfactory reduction in the $O(n^2)$ time complexity.

5. Simulation Experiment

In order to better filtered Gaussian noise in the image, we realized image denoising of the transform domain with a secondary filter way. analyzes the wavelet coefficients at different scales got by the wavelet transform of the original image, Similar block calculation is divided into two categories, one is patches matching on the boundary of two different objects, other is patches matching on the same objects.

Rough entropy important measure proposed by the previous model, the reduction transform domain coefficients of contraction method, and then the image reconstructed by the inverse transform.

We simulated this algorithm in order to verify the effectiveness and feasibility of this algorithm, the experiment takes the lena image ,Barbara image with the size of 256×256 and the grayscale of 256 as the original image, adds a normal random noise with the standard deviation value of 20 to the original image next, then conducts multi-wavelet decomposition and the denoising of wavelet coefficient's adaptive soft threshold, finally reconstructs the image. The patch size is kept square and fixed to 4×4 in all the experiments. partition the image into non-overlapping 2×2 granules. The experimental results are shown in Figure 1:



Figure 3. The Effect Picture of Lena Image Denoising

Figure 3 shows the reconstructed image of denoised lena image. Of which, Figure (a) is the original image, Figure (b) is the noise-added image, Figure (c) is the denoised image. As can be seen in Figure 1, compared with NLM denoising method, this method better preserves detail and texture information, effectively removes the noise of image, and improves the quality of image as it adopts the rough entropy attribute importance model and makes full use of information from multiple wavelet direction.

The noise residuals of the proposed method are as follows:

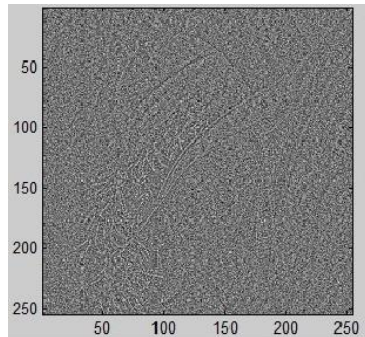


Figure 4. The Noise Figure using the Proposed Method of Filtering

As can be seen from the above noising figure, the proposed denoising method contains some structural information, but has been relatively close to the noise signal.

This paper adopts PSNR (Peak Signal to Noise Ratio) as a comprehensive evaluation index of denoising performance. The Gaussian noise with the standard deviation of 5, 10, 15, 20, 50 are added to the test images “Lena” and “Barbara”. By adopting the NL Method and our method respectively, the statistical results of denoised PSNR are shown as follows:

Table 1. The Comparison of Two Thresholds’ Denoising Method PSNR

image	method	$\sigma / PSNR$				
		5	10	15	20	50
Lena	Our method	38.05	35.43	33.73	32.46	27.85
	NLM method	37.73	35.38	33.64	32.26	27.49
Barbara	Our method	36.97	34.29	32.39	31.03	25.61
	NLM method	36.54	34.12	32.23	30.80	25.62

As can be seen from the table, after comparing the application of NLM algorithms to different images and noise, we found that it is feasible to calculate the similarity of high-frequency range of quantum by using Rough entropy attribute importance model, which prove to be having certain denoising effect.

For a visual representation, we compare the denoising results obtained with the RSM with the denoising results obtained with the NLM in the case of the noise standard deviation as follows:

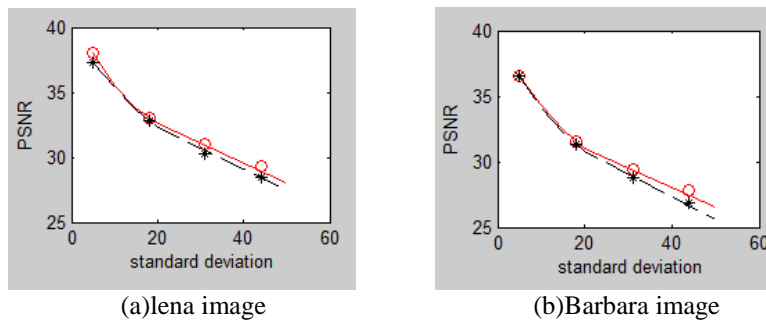


Figure 5. The Comparison of Two Thresholds’ Denoising Method PSNR

As can be seen from the figure, circle represents the RSM, while asterisk represents the NLM. The comparison of the denoising of two images “lena” and “barbana” shows that the denoising effect between the two are similar in the case of a lower standard deviation, but the corresponding PSNR of RSM is superior to that of the NLM with the increase of the standard deviation. This is due to the fact that the interference on the gray values of the image will be greater with the increase of noise intensity. If the NLM is adopted to remove noise from the image which contains strong noise, the Euclidean distance which is based on the gray value can't accurately measure the similarity between pixels, which may eventually affect the results of denoising. By using the RSM method, a better denoising effect for boundaries can be achieved due to the use of the approximate relationship of the upper and lower boundaries to distinguish between different objects.

6. Conclusion

Based on the information theory of Rough set theory, we propose a novel approach to explore image denoising for patch based image process. This paper puts forward the importance measurement model of Rough entropy and the importance reduction method of wavelet coefficients; combines rough set with thresholding method; and makes full use of lots of redundant information in natural images. On the basis of the importance model, this paper reduces those collections with high-level similarity and combines them effectively by making rational use of wavelet information from multiple directions. The simulation experiment shows that, rough set approach provides to look more in deeper to boundary patches, thus the proposed method has better peak signal-to-noise ratio, and good to protect the structure of image information.

Acknowledgements

The authors would like to thank Dr. Li Haisheng, Drs. Du Linlin for the valuable discussions. The authors are grateful to Liu Haiwen and Wang Shan for his great help in study and research. The authors are grateful to the reviewers for their thorough review and highly appreciate the comments. This work was supported by the Yong Foundation of Humanities and Social Sciences of MOE (Ministry of Education In China) of under Grant (13YJCZH089).

References

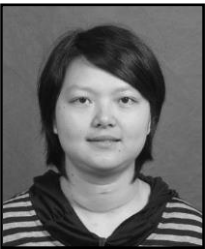
- [1] D. L. Donoho and I. M. Johnstone, “Ideal spatial adaptation by wavelet shrinkage”, *Biometrika*, vol. 81, no. 3, (1994), pp. 425-455.
- [2] D. L. Donoho, “Denoision by soft-thrsholding”, *IEEE Transactions on Information Theory*, vol. 41, no. 3, (1995), pp. 613-627.
- [3] A. Buades, B. Coll and J. M. Morel, “A non-local algorithm for image denoising”, in: *IEEE Computer Vision and Pattern Recognition*, (2005), pp. 60–65.
- [4] C. Tomasi and R. Manduchi, “Bilateral filtering for gray and color images”, in: *IEEE International Conference on Computer Vision*, (1998), pp. 839–846.
- [5] T. Brox and D. Cremers, “Iterated nonlocal means for texture restoration [A]”, In *Proc. International Conference on Scale Space and Variational Methods in Computer Vision [C]*, F Sgallari, A Murli, N Paragios, et al. New York: Springer, vol. 4485, (2007), pp. 13-24.
- [6] K. Dabov, A. Foi, V. Katkovnik and K. Egiazarian, “Image denoising by sparse 3-D transform-domain collaborative filtering”, *IEEE Trans. Image Process*, vol. 8, no. 16, (2007), August, pp. 2080-2095.
- [7] D. Lia nd Y. Du, “Artificial Intelligence with Uncertainty”, National Defense Industry Press, (2005).
- [8] W. Xu and X. Zhang, “The Rough-Entropy-Based Measurement of Uncertainty in Order Information System”, *Journal of engineering mathematics*, vol. 26, no. 2, (2009), pp. 283-290.
- [9] J. Chi and Z. Wang, “Eye Track”, China Machine Press, (2011).
- [10] D. L. Donoho and I. M. Johnstone, “Ideal denoising in an orthogonal basis chosen from a library of bases”, *C R A cad Sc-Math*, vol. 319, (1994), pp. 1317-1322.

- [11] M. Jansen, "Noise reduction by wavelet three holding", New York: Springer-Verlag, (2001).
- [12] A. Phophalia, "Rough set based image denoising for brain MR images", Signal Processing, vol. 103, (2014) October, pp. 24-35
- [13] W. Jun, H. Z. Xiao and H. R. Jiong, "Research and Application of Denoising Method Based on Wavelet Threshold", 2010ICIECS, (2010).
- [14] Y. S. Dong, "wavelet denoising based on four improved function for threshold estimation", Journal of math, vol. 26, no. 5, (2006), pp. 473-477.

Authors



Lan Song, she worked in school of Information Engineering from the East China Jiaotong University, China since 2007.6. She is working for a PH. D. degree at Computer Department of Wuhan University, Hubei province. Her research interests include wireless sensor networks, intelligent signal processing, dependable software.



Lixia Lei, she worked in school of information Engineering from the East China Jiaotong University, China. She has received the Master degree in the School of Information Engineering from the East China Jiangtong University, China. Her research interests include wireless sensor networks, intelligent signal processing.



Haisheng Li, he worked in school of Information Engineering from the East China Jiaotong University, China. He is working for a PH. D. degree at school of Computer Science and Engineering from University of Electronic Science and Technology of China. Her research interests include quantum image processing, quantum computing.



Hong Wang, he worked in Nanchang city sports bureau, Nanchang, China. He has received the Master degree in the School of eEconomics and Management from Jiangxi University of Finance and Economics. He research interests include wireless sensor networks, software engineering.