International Journal of u- and e- Service, Science and Technology
Vol.9, No. 4 (2016), pp.301-310
http://dx.doi.org/10.14257/ijunesst.2016.9.4.30

Application of Intelligent Video Monitoring System in Electric Power Construction

Yao Nan[1], Wang KaiSheng[2] and Cai Yue[3]

[1]*Department of Jiangsu Electric Power Company Research Institute, NanJing Jiangsu, 211107, China*
[2]*Department of Yangzhou Power Supply Company, YangZhou Jiangsu, 225000, China*
[3]*Department of Nanjing Yinshi Software Co.,Ltd., NanJing Jiangsu, 210037, China*
yaon_js@aliyun.com

Abstract

Since the image based intelligent video monitoring system has limited viewing angle, the blind monitoring zone will easily appear when the target is not in the field range of the camera. In order to solve the above problem, an intelligent video monitoring system with auditory function is proposed in this article on the basis of the advantages of sound localization. Firstly, the linear microphone array is acquired and the time delay estimation technology is adopted for sound localization; secondly, the camera is driven to turn to the sound source position to acquire video information; finally, the image detection program is adopted to locate and track the target in a real-time manner, and meanwhile the system feasibility is verified through the simulation experiment. The result shows that the system has good localization and tracking accuracy.

Keywords: *Sound information; Video information; Intelligent monitoring; Microphone array; Time difference of arrival (TDOA)*

1. Introduction

The intelligent monitoring system aims at, without any human intervention, automatically analyzing the acquired image sequence, locating, identifying and tracking the target in the monitoring scene so as to timely send alarms for abnormal conditions or provide valuable reference information, and such system is widely applied in security and protection systems[1].

At present, the intelligent monitoring system is mainly used to analyze and process the video images, and then corresponding software is adopted to extract the key information therein in order to identify and track the target [2-3]. Due to the limited video monitoring range, the information of the monitoring scene cannot be acquired in a full range. As a result, there are many blind zones and such video monitoring system cannot meet people's requirements[4]. In recent years, along with the gradual maturity of the microphone array signal processing technology, the sound localization system[5] based on microphone array is developed. According to the localization principle, the microphone array localization technology is mainly divided into three types: controllable beam forming technology based on maximum output power, spectrum estimation technology based on high resolution, TDOA (Time Difference of Arrival) based technology [6]. Therein, the controllable beam forming technology is sensitive to initial value and needs to know the priori knowledge regarding sound source and noise, thus having poor real-time processing property; the spectrum estimation technology based on high resolution has large

ISSN:2005-4246 IJUNESST

computation workload and cannot process the signal with high irrelevancy; TDOA based technology has small computation workload and can be easily realized, thus to become the research hotspot in recent years [7]. However, due to the existence of noise and reverberation in real sound field environment, the noise and reverberation resistance capability of TDOA based technology is not strong, and the sound localization accuracy is accordingly influenced. As a result, big error usually exists in the real-time target localization and tracking process[8]. The wavelet transform can decompose the acoustical signal including noises in different scales and then reconstruct the denoised information, thus to strengthen the acoustical signal[9]. Some scholars have introduced it into the intelligent video monitoring system, wherein sensors are used to acquire the acoustical signals in a real-time manner, and then the acoustical signals are analyzed to determine the sound source position in order to locate and track the sound source target in a real-time manner. Obviously, the monitoring range of the intelligent monitoring system can be broadened by virtue of wavelet transform.

In order to improve the monitoring effect of the intelligent monitoring system and more efficiently locate and track the target in a real-time manner, an intelligent video monitoring system with auditory function is proposed in this article. Firstly, the linear microphone array is acquired and the time delay estimation technology is adopted for sound localization; secondly, the camera is driven to turn to the sound source position to acquire video information; finally, the image detection program is adopted to locate and track the target in a real-time manner, and meanwhile the system feasibility is verified through the simulation experiment.

2. Intelligent Video Monitoring System

2.1. Hardware System

The intelligent video monitoring system mainly includes hardware system and software system, wherein the hardware system is mainly composed of processor, memorizer, camera, video encoding circuit, audio equipment, *etc*., as shown in Figure 1.

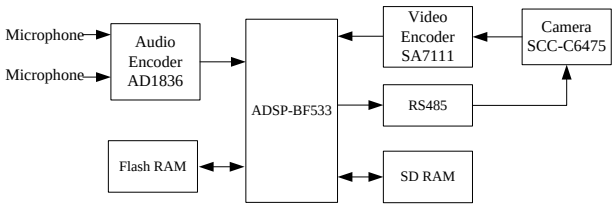

Figure 1. Hardware System Structure

2.2. Software System

2.2.1. Microphone Array Model: For a distant-field narrow-band zero-mean incident signal, unit vector $\boldsymbol{\alpha}$ denotes signal incidence direction, vector $\mathbf{p}_i$ denotes array element coordinate and the geometric structure thereof is as shown in Figure 2.

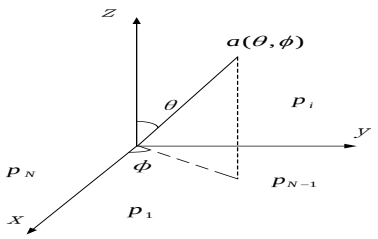

Figure 2. Geometric Structure of Array

$\boldsymbol{\alpha}$ can be expressed as:

$$\boldsymbol{\alpha} = \begin{bmatrix} -\sin\theta\cos\varphi \\ -\sin\theta\sin\varphi \\ -\cos\theta \end{bmatrix} \tag{1}$$

In the above formula, θ denotes the pitch angle and φ denotes the azimuth angle.

Array element coordinate $\mathbf{p}_i$ can be expressed as:

$$\mathbf{p}_i = \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} \tag{2}$$

In the above formula, i denotes array element number, $i = 1, 2, \ldots, N$.

The complex representation of the signal received at the origin is as follows:

$$\tilde{x}_0 = x_0(t)e^{jw_c t} + n_0(t) \tag{3}$$

In the above formula, $n_0(t)$ denotes the noise signal.

The vector representation of the signals received by the whole array is as follows:

$$\boldsymbol{x(t)} = \begin{bmatrix} x_1(t) \\ \cdot \\ \cdot \\ \cdot \\ x_N(t) \end{bmatrix} = x_0(t)e^{jw_c t} \begin{bmatrix} e^{-jw_c\tau_1} \\ \cdot \\ \cdot \\ \cdot \\ e^{-jw_c\tau_n} \end{bmatrix} + \begin{bmatrix} n_1(t) \\ \cdot \\ \cdot \\ \cdot \\ n_N(t) \end{bmatrix} \tag{4}$$

The wave number vector is defined as follows:

$$\boldsymbol{k} = -\frac{2\pi}{\lambda} \begin{bmatrix} \sin\theta\cos\varphi \\ \sin\theta\sin\varphi \\ \cos\theta \end{bmatrix} \tag{5}$$

Accordingly, the array manifold vector is expressed as follows:

$$\boldsymbol{v}(k) = \begin{bmatrix} e^{-jk^T p_1} \\ \cdots \\ e^{-jk^T p_n} \end{bmatrix} \tag{6}$$

Array signals are usually processed at the baseband and there is no any carrier component, so the received signal can be expressed as follows:

$$x(t) = x_0(t)v(k) + n_0(t) \tag{7}$$

2.2.2. Acoustical Signal Denoising: The filter coefficients of the orthogonal wavelet bases are h_{0k}, h_{1k} , the scaling function and the wavelet function are respectively $\phi(t)$ and $\psi(t)$, and the two scaling relations are as follows:

$$\begin{cases} \phi(t) = \sqrt{2} \sum_k h_{0k} \phi(2t - k) \\ \psi(t) = \sqrt{2} \sum_k h_{1k} \phi(2t - k) \end{cases} \tag{8}$$

In order to more carefully observe high-frequency components, db3 wavelet packet is adopted to decompose the acoustical signal in three classes, and the decomposition process is as shown in Figure 3.

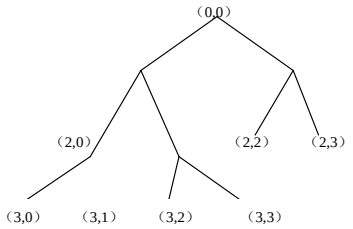

Figure 3. Wavelet Decomposition of Acoustical Signal

On the basis of setting a suitable threshold value, only the wavelet coefficients exceeding the threshold value are adopted to reconstruct the acoustical signal and eliminate noise. The threshold value is selected as follows:

$$T = c\sigma_j \tag{9}$$

In the above formula, *j* is the wavelet transform scale, and *C* is a number between 3.0 and 4.0.

$$\sigma_j = M_j / 0.6745 \tag{10}$$

In the above formula, M_j is the mean value of the absolute values of the wavelet coefficients at scale *j*.

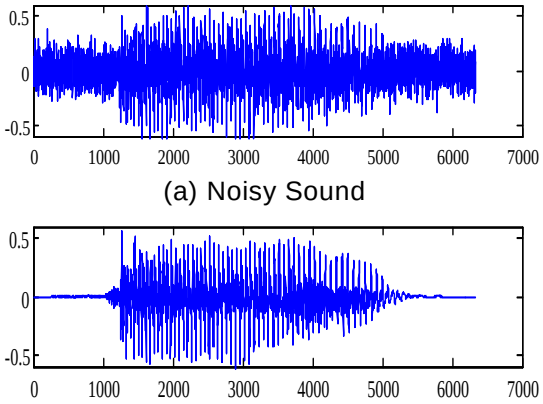

Figure 4. Wavelet Denoising Effect

According to Figure 4, for the sound denoised by wavelet transform, most noises are eliminated and meanwhile the useful sound information is reserved.

2.2.3. Acoustical Signal Endpoint Detection: Speech waveform time-domain signal is set as $x(l)$ and the n^{th} frame speech signal is set as $x_n(m)$, then $x_n(m)$ shall meet the following condition:

$$x_n(m) = w(m)x(n+m) \tag{11}$$

In the above formula, $0 \le m \le N-1$.

$$w(m) = \begin{cases} 1 & m = 0 \sim (N-1) \\ 0 & other \end{cases} \tag{12}$$

In the above formula, N is the frame length and T is the frame offset length.
The short-time energy spectrum E_n of $x_n(m)$ is defined as follows:

$$E_n = \sum_{m=0}^{N-1} [w(m)x(n+m)]^2 \tag{13}$$

Zero-crossing rate Z_n is defined as follows:

$$Z_n = \frac{1}{2N} \sum_{m=0}^{N-1} \left| \mathrm{sgn}[x_n(m)] - \mathrm{sgn}[x_n(m-1)] \right| \tag{14}$$

In the above formula, sgn[x] is defined as follows:

$$\mathrm{sgn}[x] = \begin{cases} 1 & (x \ge 0) \\ -1 & (x < 0) \end{cases} \tag{15}$$

The acoustical signal endpoint detection result is as shown in Figure 5.

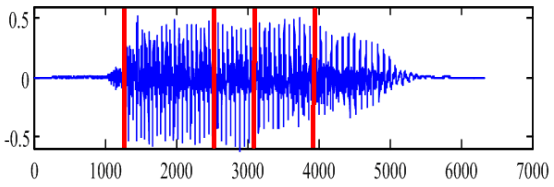

Figure 5. Acoustical Signal Endpoint Detection

2.2.4. TDOA Estimated Time Delay: If the sound source signal is set as $s(t)$, then the m^{th} acoustical signal and the n^{th} acoustical signal received by the microphone are respectively as follows:

$$x_m(t) = a_m s(t-\tau_m) + u_m(t) \tag{16}$$

$$x_n(t) = a_n s(t-\tau_n) + u_n(t) \tag{17}$$

In the above formula, $u_m(t)$ and $u_n(t)$ are additive noises, a_n and a_m are attenuation coefficients.
The time delay difference of the acoustical signal is as follows:

$$\tau = t_m - t_n \tag{18}$$

Under the condition of large PSNR, the cross-correlation function of the signals received by the two microphones is as follows:

$$R_{mn}(t) = \int_0^t s_m(t) s_n(t-\tau) d\tau \tag{19}$$

The estimated time delay of the signals received by the two microphones is as follows:

$$D = \arg\max R_{x_1,x_2}(\tau) \circ$$

2.2.5. Sound Source Position Estimation: The distance from the sound source $S(x,y,z)$ to the microphone M_i is r_i; τ_{ij} denotes the time delay from the sound source respectively to the microphone M_i and the microphone M_j; d is the distance from the microphone array element to the origin,; C is sound velocity; r, $\theta[0^\circ, 90^\circ]$ and $\varphi[0^\circ, 360^\circ]$ respectively denote the distance from the sound source to the coordinates origin, the pitch angle and the azimuth angle.

Following equation set is established according to distance and velocity:

$$\begin{cases} x^2 + y^2 + z^2 = r^2 \\ (x-d)^2 + y^2 + z^2 = r_1^2 \\ x^2 + (y-d)^2 + z^2 = r_2^2 \\ (x+d)^2 + y^2 + z^2 = r_3^2 \\ x^2 + (y+d)^2 + z^2 = r_4^2 \\ r_2 - r_1 = \tau_{12} c \\ r_3 - r_1 = \tau_{13} c \\ r_4 - r_1 = \tau_{14} c \end{cases} \tag{20}$$

After equation solution, the following solutions can be obtained:

$$\begin{cases} r \approx \dfrac{(\tau_{12}^2 + \tau_{14}^2 - \tau_{13}^2) c}{2(\tau_{13} - \tau_{12} - \tau_{14})} \\ \varphi = \arctan \dfrac{\tau_{14} - \tau_{12}}{\tau_{13}} \\ \theta = \arcsin \dfrac{c\sqrt{(\tau_{13}^2 + (\tau_{14} - \tau_{12})^2}}{2d} \end{cases} \tag{21}$$

In conclusion, the workflow of the intelligent video monitoring system with auditory function is as shown in Figure 6.

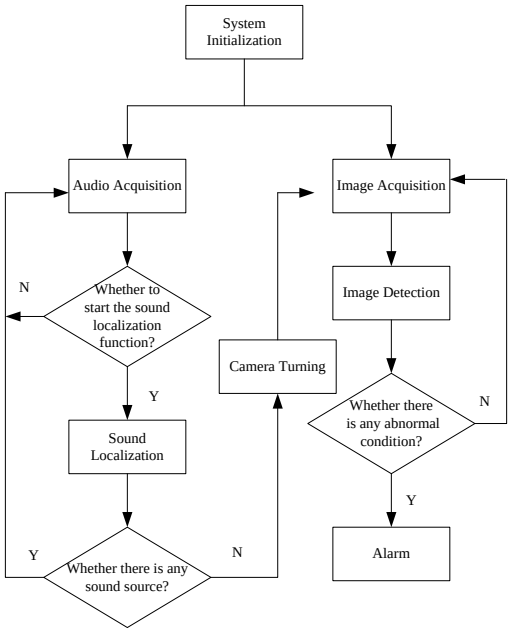

Figure 6. Workflow of Intelligent Video Monitoring System

3. Simulation Experiment

3.1. Sound Source Direction Estimation Accuracy

After the system is powered on, the addressor takes the microphone sequence as the center and then walks back and forth while taking, and then the sound will drive the camera to turn. The actual direction and the estimated direction of the sound source are as shown in Figure 7. According to Figure 7, the error between the direction estimated by the algorithm proposed in this article and the actual sound source direction is fairly small, and the result shows that the sound source estimation algorithm proposed in this article can accurately locate the target position.

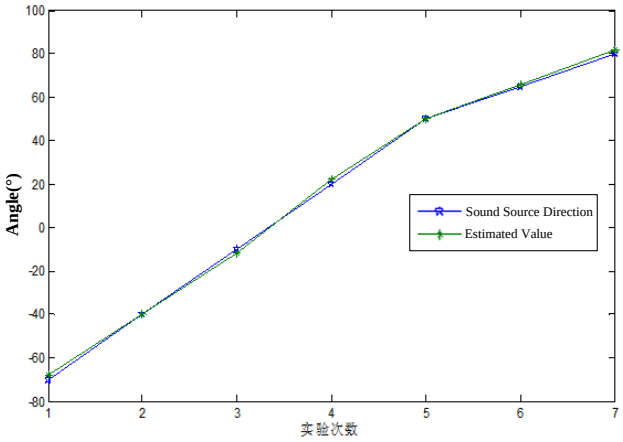

Figure 7. Angle Comparison Chart

3.2. Comparison of Sound Source Direction Estimation Accuracy before and after Denoising

The localization algorithm without wavelet-denoising process is selected for the comparison experiment, and the estimation error result is as shown in Figure8. According to Figure 8, the comparison algorithm fails to denoise the acoustical signal and accordingly has big localization error; for the algorithm proposed in this article, the wavelet transform is adopted to decompose the acoustical signal in different scales in order to eliminate the adverse impact of the noise on the acoustical signal endpoint detection and make the reconstructed acoustical signal convenient for subsequent time delay estimation, thus to strengthen the anti-noise ability of the algorithm and significantly improve the sound source localization accuracy.

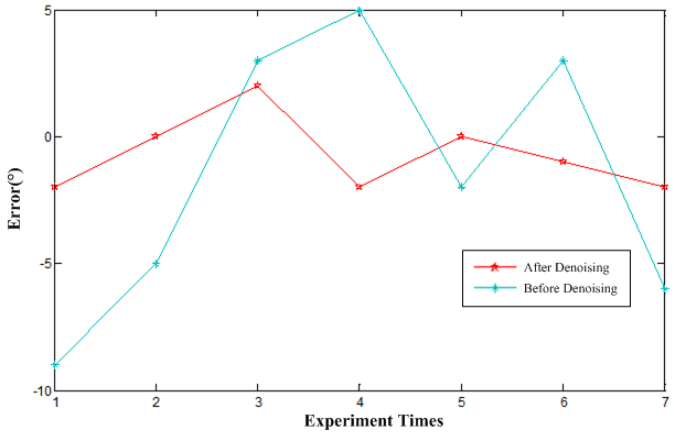

Figure 8. Comparison of Localization Accuracy Before and After Denoising

3.3. Comparison with the Tracking Performances of other Methods

The algorithms adopting single audio information and single video information are adopted to locate and track the target, and are compared with the algorithm proposed in this article, and the false tracking rate is taken as the performance evaluation standard. The results obtained thereby are as shown in Table 1.

Table 1. Comparison of False Tracking Rates of Different Algorithms

Monitoring Method	Audio Information	Video Information	Algorithm in this Article
False Tracking Rate	18.15%	10.64%	6.32%

According to Table 1, the algorithm adopting single audio information has relatively low target localization accuracy, poorest tracking effect and high false tracking rate; compared with the former algorithm, the algorithm adopting video information has improved target localization accuracy, reduced false tracking rate, more table tracking result, but still has big error; through adopting the acoustical signal to drive the camera to turn, the algorithm proposed in this article can integrate audio information and video information, thus to have better target tracking effect, lower false tracking rate and stronger robustness.

4. Conclusion

In allusion to the problem of the blind zone in present intelligent video monitoring system, an intelligent video monitoring system with auditory function is proposed in this article. Specifically, the sound localization of the microphone array is firstly adopted to detecting the target direction, and then the camera is driven to turn according to the target position. The simulation result shows that such system can improve target localization and tracking accuracy and effectively reduce false report and missing report, thus to have certain application value in modern security and protection field.

References

[1] J. Bing, X. Ke and L. Jiankun, "Analysis and management of the operation and management of power generation units during the low load period", Jiangsu electrical engineering, **(2015)**.
[2] H. Ying, "Research on transmission network planning based on improved hybrid differential evolution algorithm", Jiangsu electrical engineering, **(2015)**.
[3] S. Mingchi and Y. Dong, "Research and optimization of Nanjing electric power communication transmission network", Jiangsu electric engineering, **(2015)**.
[4] K. Wang, "Next generation job management systems for extreme-scale ensemble computing", Proceedings of the 23rd international symposium on High-performance parallel and distributed computing. ACM, **(2014)**.
[5] T. Li, "Distributed Key-Value Store on HPC and Cloud Systems", 2nd Greater Chicago Area System Research Workshop (GCASR), **(2013)**.
[6] Y. Geng, J. Chen, R. Fu, G. Bao and K. Pahlavan, "Enlighten wearable physiological monitoring systems: On-body rf characteristics based human motion classification using a support vector machine", IEEE transactions on mobile computing, vol. 1, no. 1, **(2015)** April, pp. 1-15.
[7] J. He, Y. Geng, F. Liu and C. Xu, "CC-KF: Enhanced TOA Performance in Multipath and NLOS Indoor Extreme Environment", IEEE Sensor Journal, vol. 14, no. 11, **(2014)** November, pp. 3766-3774.
[8] S. Zhou, L. Mi, H. Chen and Y. Geng, "Building detection in Digital surface model", 2013 IEEE International Conference on Imaging Systems and Techniques (IST), **(2012)** October.
[9] J. He, Y. Geng and K. Pahlavan, "Toward Accurate Human Tracking: Modeling Time-of-Arrival for Wireless Wearable Sensors in Multipath Environment", IEEE Sensor Journal, vol. 14, no. 11, **(2014)** November, pp. 3996-4006.
[10] N. Lu, C. Lu, Z. Yang and Y. Geng, "Modeling Framework for Mining Lifecycle Management", Journal of Networks, vol. 9, no. 3, **(2014)** January, pp. 719-725.
[11] Y. Geng and K. Pahlavan, "On the accuracy of rf and image processing based hybrid localization for wireless capsule endoscopy", IEEE Wireless Communications and Networking Conference (WCNC), **(2015)** March.
[12] G. Liu, Y. Geng and K. Pahlavan, "Effects of calibration RFID tags on performance of inertial navigation in indoor environment", 2015 International Conference on Computing, Networking and Communications (ICNC), **(2015)** Febuary.
[13] J. He, Y. Geng, Y. Wan, S. Li and K. Pahlavan, "A cyber physical test-bed for virtualization of RF access environment for body sensor network", IEEE Sensor Journal, vol. 13, no. 10, **(2013)** October, pp. 3826-3836.
[14] W. Huang and Y. Geng, "Identification Method of Attack Path Based on Immune Intrusion Detection", Journal of Networks, vol. 9, no. 4, January **(2014)**, pp. 964-971.
[15] G. Bao, L. Mi, Y. Geng, M. Zhou and K. Pahlavan, "A video-based speed estimation technique for localizing the wireless capsule endoscope inside gastrointestinal tract", 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), **(2014)** August.
[16] D. Zeng and Y. Geng, "Content distribution mechanism in mobile P2P network", Journal of Networks, vol. 9, no. 5, **(2014)** January, pp. 1229-1236.
[17] M. Zhou, G. Bao, Y. Geng, B. Alkandari and X. Li, "Polyp detection and radius measurement in small intestine using video capsule endoscopy", 2014 7th International Conference on Biomedical Engineering and Informatics (BMEI), **(2014)** October.
[18] G. Yan, Y. Lv, Q. Wang and Y. Geng, "Routing algorithm based on delay rate in wireless cognitive radio network", Journal of Networks, vol. 9, no. 4, Jan. **(2014)**, pp. 948-955.
[19] G. Bao, L. Mi, Y. Geng and K. Pahlavan, "A computer vision based speed estimation technique for localiz ing the wireless capsule endoscope inside small intestine", 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), August **(2014)**.
[20] Y. Lin, J. Yang, Z. Lv, W. Wei and H. Song, "A Self-Assessment Stereo Capture Model Applicable to the Internet of Things", Sensors, **(2015)**.

[21] W. Ou, Z. Lv and Z. Xie, "Spatially Regularized Latent topic Model for Simultaneous object discovery and segmentation", The 2015 IEEE International Conference on Systems, Man, and Cybernetics SMC, **(2015)**.
[22] K. Wang, "Using Simulation to Explore Distributed Key-Value Stores for Exascale System Services", 2nd Greater Chicago Area System Research Workshop (GCASR), **(2013)**.
[23] Y. Wang, Y. Su and G. Agrawal, "A Novel Approach for Approximate Aggregations Over Arrays", In Proceedings of the 27th international conference on scientific and statistical database management, ACM, **(2015)**.
[24] Z. Lv, A. Halawani, S. Feng, S. ur Rehman and H. Li, "Touch-less Interactive Augmented Reality Game on Vision Based Wearable Device", Personal and Ubiquitous Computing, **(2015)**.
[25] J. Yang, S. He, Y. Lin and Z. Lv, "Multimedia cloud transmission and storage system based on internet of things", Multimedia Tools and Applications, **(2016)**.
[26] X. Song and Y. Geng, "Distributed community detection optimization algorithm for complex networks", Journal of Networks, vol. 9, no. 10, **(2014)** January, pp. 2758-2765.
[27] J. Hu and Z. Gao, "Distinction immune genes of hepatitis-induced heptatocellular carcinoma", Bioinformatics, vol. 28, no. 24, **(2012)**, pp. 3191-3194.
[28] J. Hu, Z. Gao and W. Pan, "Multiangle Social Network Recommendation Algorithms and Similarity Network Evaluation", Journal of Applied Mathematics, **(2013)**.

Author

YaoNan was born in Jiangsu, China, in 1976. He Received M.S. Degrees in Computer software engineering from NanJing University, JiangSu, China, in 2005, respectively. He is currently work in Jiangsu Electric Power Company Research Institute. His research direction includes the power image intelligent technologyand information technology. He has received a number of scientific and Technological Progress Award.

