TELKOMNIKA Indonesian Journal of Electrical Engineering
Vol.12, No.5, May 2014, pp. 3902 ~ 3910
DOI: http://dx.doi.org/10.11591/telkomnika.v12i5.4816

Acoustic Source Localization Based on Iterative Unscented Particle Filter

Jie Cao, Jia-qi Liu*, Di Wu, Jin-hua Wang
College of Electrical and Information Engineering, Lanzhou University of Technology,
Lanzhou 730050, P.R. China
*Corresponding author, e-mail: 450971614@qq.com

Abstract

To solve the problem of tracking an acoustic source in noise and reverberation environment, a new method is proposed in pursuit of higher accuracy. First, this paper improves the unscented particle filter, which can add the latest measurement information to optimize the proposal distribution. Then, the likelihood function is constructed by calculating the microphone arrays' output energy in the framework of the improved algorithm. Finally, the experiment results indicate that the proposed localization method can not only improves the accuracy of location estimation, but also can enhance the ability to resist noise and reverberation in the acoustic source localization system.

Keywords: *computer application, acoustic source localization, iterative unscented particle filter (IUPF), microphone arrays, proposal distribution function*

1. Introduction

Speaker localization based on microphone array is an important topic in Human-Computer Interaction Research. It has widely applications in several fields, including the multimedia systems, video conferencing systems and mobile robotics and so on [1]. The traditional speaker localization based on microphone array mainly estimated the speaker location by calculating the current time delay of the speech signal received by microphone array. In the situation of free sound field, this method can get a great effect of locating and tracking. However, it will generate lots of fake sound resource under a strong background noise or the situation of long reverberation time. As a result, it will affect the accuracy of localization.

In recent years, with the development of non-linear filtering techniques, researcher modeled the speaker movement tracking using the way of state-space method, simulating the situation of speaker's movement with proper dynamical equation, synthesizing present information with the past ones, surmount the effect of virtual sound source effectively in complex noise situation, and improve the accuracy and robustness of speaker localization system. Vermaak [2] introduced the particle filter algorithm to speaker localization system, established a reasonable speaker motion model to suppress spurious noise sources, and constructed the likelihood function according to the time delay estimation. Ward [3] made improvements on the basis of the previous, summarized the sound source locate methods and used output energy of a steered beam-former to construct the likelihood function, achieving accurately tracking of speaker by using the particle filter. Respectively, on the basis of improved particle filter, Fu-Liang Yin [4, 5] constructed likelihood function based on time-delay estimation and the output energy of SRP-PHAT beamformer for tracking and locating the speaker. Nai-gao Jing [6] used quantum evolutionary methods to improve particle filter and applied the improved particle filter to speaker localization, and achieved good results. The above methods can accurately locate the speaker in short reverberation time and large SNR environment, but still cannot effectively locate the speaker in the strong noise environment.

Therefore, this paper presents an acoustic source localization method based on iterative unscented particle filtering method, which considers the inhibition room reverberation role of measurement information and speaker motion model aiming to the inaccurate problem of speaker location in the environment of long reverberation time and small SNR. The method uses the audio signal collected by the microphone array as the observation information, and

Received October 20, 2013; Revised December 5, 2013; Accepted December 31, 2013

constructs the likelihood function by calculating the steered beamformer energy formed by the microphone array. On the other hand, the introduction of iteration unscented particle filter and the analysis of comparing to other filtering algorithms verified the effectiveness of this algorithm in acoustic source localization.

In the rest of this paper, we explain our algorithm in Section 2. Experimental results and analysis are reported in Section 3. We conclude this paper in Section 4.

2. The Proposed Algorithm

2.1. Particle Filter

Particle filter is a filtering method based on Monte Carlo and recursive Bayesian estimation. In recent years, it has become a common tool for locating target under non-linear or non-Gaussian conditions [7,8,9]. The core idea is that using the weighted sum of a series of random samples to represent the posterior probability density. Assuming that nonlinear system dynamic model as follows:

State equation: $x_k = f\left(x_{k-1}\right) + v_{k-1}$ (1)

Measurement equation: $z_k = h\left(x_k\right) + n_k$ (2)

Where x_k is the system state, z_k is the observed state, the map $f(\cdot)$ and $h(\cdot)$ represent the system state transition model function and measurement model function, v_{k-1} and u_k are the process noise and observation noise.

Let $\left\{x_k^i, w_k^i, i = 1, \cdots, N\right\}$ represents a set of random weighted sample (particles), where x_k^i is the i -th particle state in time k , the corresponding weight value is w_k^i , there:

$$p(x_k \left| Y_k \right.) = \sum_{i=1}^{N} w_k^i \delta(x - x_k^i) \quad (3)$$

$$w_k^{(i)} \propto w_{k\text{-}1}^{(i)} \frac{p(Y_k \mid x_k^{(i)}) p(x_k^{(i)} \mid x_{k-1}^{(i)})}{q(x_k^{(i)} \mid x_{0:k-1}^{(i)}, Z_{1:k})} \quad (4)$$

Where $\delta(x - x_k)$ is the unit impulse function, that is $\delta(x - x_k) = 0, x \neq x_k$ and $\int \delta(x) d_x = 1$, particle sets sampled from proposal distribution $q(x_k^{(i)} \mid x_{0:k-1}^{(i)}, Z_{1:k})$, whose weights satisfy the normalization condition $\sum_{i=1}^{N} w_k^i = 1$.

2.2. The Proposed Algorithm

In the acoustic source localization system, particle filter algorithm selects prior probability density as proposal distribution, this approach lost measured value of the current time, which leads to the wrong localization. While unscented particle filter (UPF) algorithm uses the unscented Kalman filter to generate proposal distribution, which can be well integrated into the latest measurement information to improve location accuracy. The detailed description refer paper [10-12], the UPF's principle will not be introduced in this section. But UPF also can not accurately locate the speaker when the filtering estimation precision is lower. This paper improves the UPF algorithm by joining iterative Kalman filter (IKF) algorithm to revise the state mean and variance updated by UKF, and optimizes the proposal distribution. IKF [13] is a kind of maximum posterior approach, which aims to find more information about using the known state estimation approximation of $\hat{x}^+$ and P^+. For any natural Number i determined by prior or convergence criteria, the measurement update formula is:

$\hat{x}^+ = x_{i+1}, P^+ = P_{i+1}$.Sequence $\{x_i\}$ and $\{P_i\}$ are defined as:

$$x_0 = \hat{x}, P_0 = P$$
$$x_{i+1} = \hat{x} + K_i(z - h(x_i) - H_i(\hat{x} - x_i)) \tag{5}$$

$$P_{i+1} = (I - K_i H_i)P \tag{6}$$

Where $K_i = PH_i^{\mathrm{T}}(H_i PH_i^{\mathrm{T}} + R)^{-1}, H_i = h'(x_i)$, R is the measurement noise covariance matrix, I is a unit matrix. It can get new mean and variance by iterative update. Combined the iterative thought of formula (5) and (6) with UPF can get iterated unscented Kalman filter (IUPF) algorithm. Assume that the system as formula (1) and (2), the process noise and measurement noise are all zero-mean Gaussian white noise which are not related to each other, covariance are Q_k and R_k, The improved part of the algorithm is described as follows:

(1) Initialization. At the time $k = 0$, generate samples from the prior distribution $\{x_0^i, i = 1, 2, \cdots, N\}$.

$$x_{0|0}^i = E(x_0^i) \tag{7}$$

$$P_{0|0}^i = E[(x_0^i - \overline{x}_0^i)(x_0^i - \overline{x}_0^i)^T] \tag{8}$$

(2) Important sampling

At time $x_{0,k-1}^i = \overline{x}_{k-1}^i$, $i = 1, 2, \cdots, N_s$, use the UKF algorithm to update the particle.

Select particle $x_{0,k-1}^i = x_{k-1|k-1}^i$.

$$x_{j,k-1}^i = x_{k-1|k-1}^i + \sqrt{(n_x + \lambda)P_{k-1|k-1}^i}, j = 1, 2, \cdots, n_x \tag{9}$$

$$x_{j,k-1}^i = x_{k-1|k-1}^i - \sqrt{(n_x + \lambda)P_{k-1|k-1}^i}, j = n_x + 1, n_x + 2, \cdots, 2n_x \tag{10}$$

$$W_0^m = \frac{\lambda}{n_x + \lambda} \tag{11}$$

$$W_0^c = W_0^m + (1 + \alpha^2 + \beta) \tag{12}$$

$$W_j^m = W_j^c = \frac{1}{2(n_x + \lambda)} \tag{13}$$

Where $W_j^{(m)}$ and $W_j^{(c)}$ are the weights coefficient of the first-order statistical properties and the second-order statistical properties.

Time Update:

$$x_{j,k|k-1}^i = f(x_{j,k-1}^i) \tag{14}$$

$$\bar{x}^i_{k|k-1} = \sum_{j=0}^{2n_x} W_j^{(m)} x^i_{j,k|k-1} \tag{15}$$

$$P^i_{k|k-1} = \sum_{j=0}^{2n_x} W_j^{(c)} \left[x^i_{j,k|k-1} - \bar{x}^i_{k|k-1} \right] \left[x^i_{j,k|k-1} - \bar{x}^i_{k|k-1} \right]^T + Q_{k-1} \tag{16}$$

$$z^i_{j,k|k-1} = h(x^i_{j,k|k-1}) \tag{17}$$

$$\bar{z}^i_{k|k-1} = \sum_{j=0}^{2n_x} W_j^{(m)} z^i_{j,k|k-1} \tag{18}$$

Measurement update:

$$P_{xz} = \sum_{j=0}^{2n_x} W_j^{(c)} \left[x^i_{j,k|k-1} - \bar{x}^i_{k|k-1} \right] \left[z^i_{j,k|k-1} - \bar{z}^i_{k|k-1} \right]^{\mathrm{T}} \tag{19}$$

$$P_{zz} = \sum_{j=0}^{2n_x} W_i^{(c)} \left[z^i_{j,k|k-1} - \bar{z}^i_{k|k-1} \right] \left[z^i_{j,k|k-1} - \bar{z}^i_{k|k-1} \right]^{\mathrm{T}} \tag{20}$$

$$K_k = P_{xz} P_{zz}^{-1} \tag{21}$$

$$x^i_{k|k} = \bar{x}^i_{k|k-1} + K_k (z_k - \bar{z}^i_{k|k-1}) \tag{22}$$

$$P^i_{k|k} = P^i_{k|k-1} + K_k P_{zz} K_k^T \tag{23}$$

(3) Iterative update

Use IKF to revise the state mean and variance updated by measurements, get new mean $\tilde{x}^i_k$ and variance $\tilde{P}^i_k$.

Particle sampling: $\hat{x}^i_k \sim q(\tilde{x}^i_k \,|\, \tilde{x}^i_{0:k-1}, z_{1:k}) = N(\tilde{x}^i_k, \tilde{P}^i_k)$ (24)

Where $N(\cdot)$ represents a Gaussian function.

Weights calculating: $w^i_k \propto \dfrac{p(z_k \,|\, \hat{x}^i_k) \cdot p(\hat{x}^i_k \,|\, x^i_{k-1})}{q(\hat{x}^i_k \,|\, x^i_{0:k-1}, z_{1:k})}$ (25)

Where $i = 1, 2, \cdots, N_s$, Normalized weights:

$$w^i_k = \frac{w^i_k}{\sum_{i=1}^{N_s} w^i_k} \tag{26}$$

(4) Resample process.

(5) State estimation:

$$x_k = \sum_{i=1}^{N_s} w_k^i x_k^i \qquad (27)$$

IUPF fully integrated into the current measurement of the latest information, can accurately locate the speaker.

3. Acoustic Source Localization Based on IUPF

3.1. Speaker Localization Function

SRP-PHAT sound source localization algorithm combined robustness of beamform method and short analysis of characteristics with the insensitivity to the environment of phase transformation method, reducing the acoustic source localization system's sensitivity to noise and reverberation, improving the system's robustness and positioning accuracy [14]. The algorithm searches the maximum output energy of the beam in the total space, to determine the sound source location, use a shorter data analysis, so this method is very suitable for real-time tracking the movement of the speaker. The specific calculations form as below.

Assuming that the sound source signal $s(t)$ reaches the microphone array through the multipath propagation. The i-th microphone received signal $m_i(t)$ can be expressed as:

$$m_i(t) = s(t) * h_i(t) + v_i(t) \qquad (28)$$

Where $v_i(t)$ is the noise, $h_i(t)$ is the impulse response between the source and the first microphone, it is a function of the microphone and the sound source location. "*"is the convolution operator. Set the space vector of the sound source is χ ,the Fourier transform of speech signal $m_i(t)$ is $M(\omega)$,the number of microphones is M ,the propagation delay of sound signal to the i-th microphone in the direct path is τ_i .SRP-PHAT sound source localization algorithm is by calculating the microphone array beam output energy to positioning, localization function is as follows:

$$y_t^{SRP}(\chi) = \sum_{i=1}^{M} \sum_{j=1}^{M} \int_{-\infty}^{+\infty} \frac{M_i(\omega) M^*_i(\omega)}{\left| M_i(\omega) M^*_i(\omega) \right|} e^{j\omega(\tau_i - \tau_j)} d\omega \qquad (29)$$

3.2. The Implementation of Proposed Algorithm

This paper use the audio signal captured by microphone arrays as observational information to locate the speaker. The IUPF algorithm framework needs to establish the motion model of the speaker and the speaker localization function. This paper use the Langevin process to establish the speaker's motion model, the detailed description refers paper [15]. The implementation steps of speaker localization based on iteration unscented particle filter as follows,

Step 1: Initialization: k=1, initialize particle sets $\left\{ x_0^i, 1/N, i = 1, \cdots, N \right\}$.

Step 2: Particles sample. $k = 1, 2, \cdots$, obtain the sampling particles at time k by using IUPF algorithm and speaker motion model and localization function.

$$\hat{x}_k^i \sim q(x_{k|k}^i \left| x_{0:k-1}^i, z_{1:k} \right.) = N(x_{k|k}^i, P_{k|k}^i) \qquad (30)$$

Step 3: Weights update. Obtain particle weights w_k^i at time k according to formula (25) and (26).

Step 4: Resample process.

Step 5: State output. Finally, particle sets and corresponding weights can be obtained. State estimation is $x_k = \sum_{i=1}^{N_s} w_k^i x_k^i$.

Step 6: Determine whether the speech signal gets an end. If true stops running, otherwise go to step 2.

4. Experimental Results and Analysis

4.1. Experimental Parameters Settings

The performance of the proposed acoustic source localization algorithm are evaluated in a simulated ($5m \times 7m \times 3m$) rectangular room, as shown in Figure 1. In two directions of X and Y, the two groups of linear arrays which containing two microphones are located respectively, and the distance between the microphones in each group is 1m, the speaker moves uniformly along the x axis inclined 45° direction in the room, and keeps speaking during the movement, and the starting point of the movement is (1,1). The reverberation impulse response function in the room is generated by the IMAGE model, using Gaussian white noise, the speech signal is obtained by microphone array sampled with the sampling rate of fs=16 KHZ. The height of the speaker is set to a constant value.

During the experiment, The parameters in the speaker movement model are $\beta_x = 10\text{s}^{-1}$, $v_x = 1\text{ms}^{-1}$, $\Delta T = 32\text{ms}$, the initial state of the speaker is $a_0 = [1, 0, 1, 0]^{\mathrm{T}}$, with the covariance of $\mathrm{P}_0 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \times \beta^2$,the transform length of FFT is L=512, window function is hamming window. Sound source moving uniformly along the x-axis inclined 45° direction with the speed of $0.1m/s$. To verify the performance of IUPF algorithm on the speaker localization, this paper compare it with the PF and the UPF which used in paper [10], and introduce the root mean square error (RMSE) as a standard of precision measurement. The root mean square error (RMSE) is defined as:

$$RMSE = \left(\frac{1}{T}\sum_{t=1}^{T}(x_k - x_k^i)^2\right)^{1/2}$$

The smaller the RMSE, the higher the position accuracy.

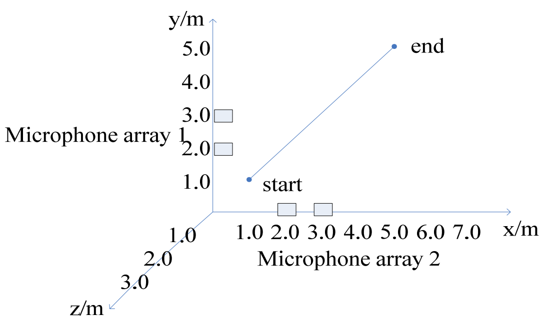

Figure 1. The Placement of Microphone Array

4.2. The Experimental Results and Analysis

Under the condition of different SNR and reverberation time (T_{60}), we compared the performance of PF, the UPF and the proposed method in this paper. Figure 2 shows the localization effect of three algorithms in X direction when $SNR = 15dB,\ T_{60} = 100ms$. Figure 3 shows the localization effect of three algorithms in Y direction when $SNR = 15dB,\ T_{60} = 100ms$. Figure 4 shows the localization effect of three algorithms in X direction when $SNR = 5dB, T_{60} = 200ms$. Figure 5 shows the localization effect of three algorithms in Y direction when $SNR = 5dB, T_{60} = 200ms$. The horizontal axis represents time, the vertical axis represents the position in each direction. Experiment has 50 times simulations respectively in two different SNR and reverberation time (T_{60}), getting the average values of the RMSE as shown in Table 1 and Table 2.

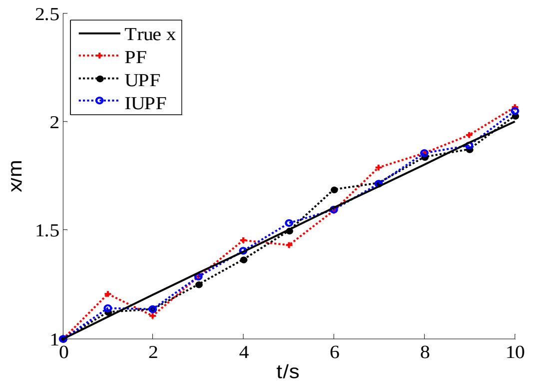

Figure 2. $SNR = 15dB, T_{60} = 100ms$, the Location Results of Three Algorithms in X Direction

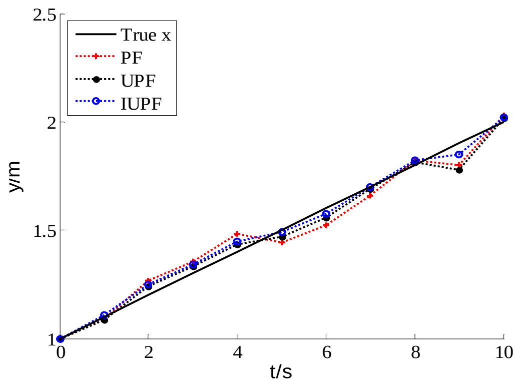

Figure 3. $SNR = 15dB, T_{60} = 100ms$, the Location Results of Three Algorithms in Y Direction

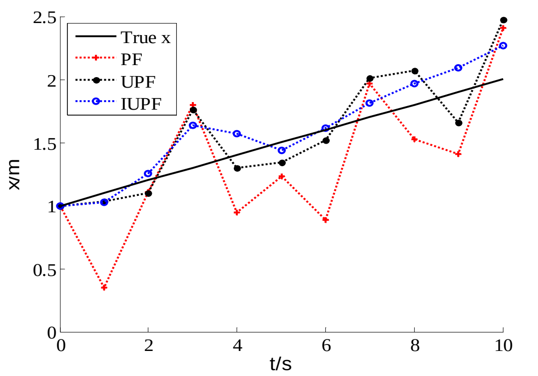

Figure 4. $SNR = 5dB, T_{60} = 200ms$, the Location Results of Three Algorithms in X Direction

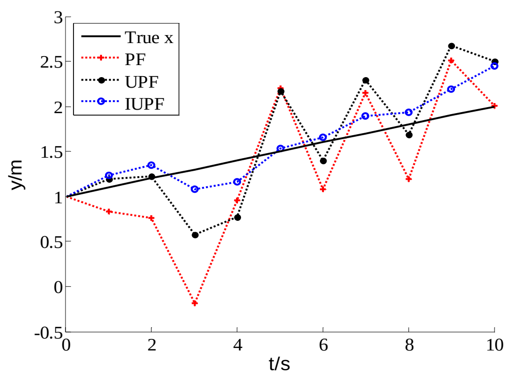

Figure 5. $SNR = 5dB, T_{60} = 200ms$, the Location Results of Three Algorithms in Y Direction

Figure 2 and Figure 3 show the location results of the PF, the UPF used in paper [10] and the proposed algorithm under the condition of higher signal noise ratio (SNR) and shorter reverberation time. Obviously, the location results of PF and UPF are coarse and imprecise. The proposed algorithm can achieve precise localization of the speaker compared to PF and UPF. Results indicate that the proposed method is superior to PF and UPF algorithm on the location accuracy in high SNR.

Figure 4 and Figure 5 show the location results of three algorithms as the SNR decreases and reverberation time increases. The location accuracy of three algorithms reduced in different degree, the estimation error of PF and UPF algorithm increased obviously while the proposed algorithm can still keep good localization accuracy. The experiments show that the stability of the proposed method is better than PF and UPF algorithms.

Table 1. $SNR = 15dB, T_{60} = 100ms$, Comparison of RMSE Value

Filtering algorithm	RMSE	
	X	Y
PF	0.0627	0.0578
UPF	0.0415	0.0451
Our algorithm	0.0333	0.0318

Table 2. $SNR = 5dB, T_{60} = 200ms$, Comparison of RMSE Value

Filtering algorithm	RMSE	
	X	Y
PF	0.4429	0.6294
UPF	0.2570	0.4906
Our algorithm	0.1655	0.2119

By comparing the mean square error of Table1 and Table 2, the IUPF algorithm has minimum RMSE, the localization accuracy is 50%-60% higher than that of the PF, 10%-30% higher than that of the UPF. It shows that the proposed algorithm in this paper is always better than the algorithm in paper [10] and the standard filter method.

In summary, under the condition of lower SNR and stronger reverberation, PF and UPF algorithm can appear large localization errors. Obviously, our algorithm greatly improves the location accuracy.

5. Conclusion

Analyzing the traditional localization method based on particle filter algorithm, an effective localization algorithm is proposed in this paper which combines UKF and IKF to generate the proposal distribution function of the PF. Under the framework of the improved algorithm, this paper constructed the likelihood function using SRP-PHAT. Simulation results show that the localization accuracy of the proposed algorithm has obvious improvement comparing to PF algorithm and UPF algorithm.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (61263031). The Finance Department Foundation of Gansu Province, China (0914ZTB148).

References

[1] Jin NG, Yin FL, Chen Z. Weighted subspace fitting sound source localization method based on particle filtering. *Journal of Electronics & Information Technology*. 2008; 30(9): 2134-2137.

[2] J Vermaak, A Blake. *Nonlinear filtering for speaker tracking in noisy and reverberant e- nvironments*. Proc. IEEE Int. Conf. Acoust. Speech, Signal Processing (ICASSP-01), Salt Lake City, UT. 2001.

[3] DB Ward, RC Williamson. *Particle filter beamforming for acoustic source localization in a reverberant environment.* Proc.IEEE Int. Conf. Acoustic, Speach, Signal Processing (ICASSP-02), Orlando, FL. 2003.

[4] Hou DW, Yin FL. An IMM particle filtering method for speaker tracking. *ACTA Electronica Sinica*. 2010; 38(4): 835-841.

[5] Jin NG, Yin FL, Chen Z. Microphone array speaker tracking using layered sampling particle filtering. *ACTA Electronica Sinica*. 2008; 36(1): 194-198.

[6] Jin NG, Yin FL. Quantum Evolutionary Particle Filtering Algorithm with Application to Speaker Tracking. *Signal Processing*. 2008; 24(6): 982-987.

[7] Gordon N, Salmond D, Smith A. *Novel approach to nonlinear/non-Gaussian Bayesian state estimation.* Proceeding of Institute Electric Engineering. 1993; 140(2): 107-113.

[8] Gao Y, Yuan GW, Zhou H, et al. Object Descriptor Combining Color and Detection. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2013; 11(6): 3072-3079.
[9] Indah A, Siradjuddin M, Rahmat W, T Basaruddin. Particle Filter with Gaussian Weighting for Human Tracking. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2012; 10(4): 801-806.
[10] ZHANG MH, LIU XX. Maneuverable target tracking in dynamic scene based on unscented particle filter. *Journal of Optoelectronics·Laser*. 2010; 6(21): 924-929.
[11] PAN Q, YANG F, YE L. Survey of a kind of nonlinear filters-UKF. *Control and Decision*. 2005; 20(5): 481-489.
[12] QU YW, ZHANG EH, YANG JY. Improved unscented particle filter .*Control Theory & Applications*. 2010; 9(27): 1152-1158.
[13] Shmaliy YS. An Iterative Kalman-Like Algorithm Ignoring Noise and Initial Conditions. *IEEE Transactions on Signal Processing.* 2011; 59(6): 2465 – 2473.
[14] Dante A Blauth ,Vicente P Minotto, Claudio R Jung, Bowon Lee, Ton Kalker. Voice activity detection and speaker localization using audiovisual cues. *Pattern recognition letters*. 2012; 33(4): 373-380.
[15] Hou DW, Yin FL,Tao Z. Quasi-Monte Carlo filtering for speaker tracking. *ACTA Automatica Sinica.* 2009; 35 (7): 1016-1021.