The impulse sound source tracking using Kalman filter and the cross-correlation
Yun Woo-han\textsuperscript{1,} Oh Cheon-in\textsuperscript{2}, Ban Kyu-Dae\textsuperscript{3} and Ji Su-young\textsuperscript{4}
\textsuperscript{1,4} Intelligent Robot Research Division, ETRI, Daejeon, Korea
(Tel : +82-42-860-{5804,5337}; E-mail: {yochin, chisy}@etri.re.kr)
\textsuperscript{2,3} Department of Computer Software & Engineering, UST, Daejeon, Korea
(Tel : +82-42-860-{1471,5925}; E-mail: {oci64380, kdban}@etri.re.kr)

Abstract: In case of robot interacting with human, natural movement and action of the robot give the feeling of closeness to people. One of the basis technique of human-robot interaction is sound source tracking. In this paper, we propose a weighted kalman filter using cross-correlation value based on TDOA method for sound source tracking. In the experiment, we use the footstep sound as an impulse sound. The experiment shows that the proposed method gives more accurate and natural azimuth results.

Keywords: Sound source tracking, Robot, Kalman filter, Cross-correlation.

1. INTRODUCTION

When the sound was made, a human being or an animal knows where the sound comes from and determines the next action. The sound source tracking technique provides a robot with a clue to determine next action and be familiar with human. In the past few years, many researchers have developed sound source tracking method\cite{1}\cite{2}\cite{3}\cite{4}.

The method for estimating an azimuth of a sound source is divided into two classes, those are based on ILD, and those are based on ITD. The method based on Interaural Level Difference (ILD) estimates the azimuth of a sound source using the decreasing rate of amplitude of sound with distance\cite{5}\cite{6}. The method based on the Interaural Time Difference (ITD) estimates the azimuth of a sound source using TDOA (Time Difference of Arrival). Generally the method based on TDOA is widely used and better than the method based on ILD through the many experiments\cite{7}\cite{8}\cite{9}. Recently the method for sound source tracking is developed based on TDOA.

When a system with just two microphones which have short distance each other estimates the azimuth using the method based on TDOA, the discretization problem of an azimuth is occurred. If we assume that the velocity of the sound is $v$ m/s, the distance between two microphones is $k$ m and the sampling rate is $x$ Hz, then the maximum of the frame difference of two microphones is $kx/v$ frames. So, the azimuth space of sound source is discretized and reduced into $2kx/v+1$ states. The sound source tracking method cannot find the accurate azimuth of the sound source. Also, the robot action looking for the sound source is not natural, because the azimuth is already not accurate.

We solve this problem using a weighted Kalman filter using cross-correlation value based on TDOA method.

The paper is organized as follow: in Section 2 we present the overall system and main procedure. The results in Section 3 shows the experiment. In Section 4, we conclude this paper.

2. TRACKING SYSTEM

2.1 System

We used the iCat made by Phillips International, Inc\cite{11}. Each microphone was detached to each foot of the robot and depart in 22 cm. Because we use two microphones, we use only the front of a robot as a test region.

![Fig. 1 The robot, named iCat, was used in the experiment.](image)

2.2 Feature detection in an input signal

To find the impulse sound in input sound stream, we calculate energy of sound based on the window and compare it with the predefined threshold. If the energy is over the threshold, we do the next process. Let $f(t)$ is the amplitude of sound at time $t$. The energy of each sound frame is calculated like this.

$$E = f(t)^2$$

We set the threshold by experiment in advance.

2.3 Cross-correlation between two microphones

When we find the impulse starting point in one microphone, we fine the corresponding impulse starting point in other microphone late. When the speed of sound is $v$ m/s, the distance between two microphones is $k$ m and the sampling rate is $x$ Hz, then the maximum delay between two microphones is $kx/v$ frames. So, we
just calculate the cross-correlation value from 0 to kx/v frames and find the peak value in them to obtain the corresponding point. Let x(t) and y(t) be the amplitude of two sound signals, respectively. The cross-correlation at time m $R_{xy}(m)$ is calculated like this

$$R_{xy}(m) = \sum_{t=0}^{N-m-1} x(t + m)y(t)$$  \hspace{1cm} (2)

### 2.4 Kalman filter

Kalman filter is very famous algorithm for tracking the object. When we obtain the observed position information of an object, we model the system with the position, velocity and acceleration. And we predict the next position of an object and update the real position using next observed position information of an object. Kalman filter algorithm is working as follow. Let z is an observed factor and x is a hidden factor.

The system has a linear relation with time.

$$x_k = Ax_{k-1} + w_{k-1}$$  \hspace{1cm} (3)

In this equation, the w, k and A means the observed noise, time k and a system matrix, respectively. The hidden factor x and the observed factor z have a linear relation.

$$z_k = Hx_k + v_k$$  \hspace{1cm} (4)

In this equation, H means an observation matrix of system. The hidden factor x is updated using the observed factor z.

$$K_k = P_k^{-1}H^T(HP_k^{-1}H^T + R)^{-1}$$  \hspace{1cm} (5)

$$x_k = x_k^- + K_k(z_k - Hx_k^-)$$  \hspace{1cm} (6)

We predict the next position x using the updated value and predict the next value z using the x.

$$x_k^- = Ax_{k-1} + Bu_{k-1}$$  \hspace{1cm} (7)

In our case, we use the azimuth calculated from the sound signal based on TDOA as the factor z and use the real azimuth and velocity as the factor x. For this model, we use the system matrix A and observation matrix H like this.

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

$$H = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

### 2.5 Estimation of an azimuth of the sound source

Finally, we estimate the actual azimuth using cross-correlation value based on TDOA and Kalman filter. The cross-correlation value is in zero to one. The large cross-correlation value means that two signals are similar and believable. But small cross-correlation value means two signals are not similar. Let $a_o(t)$ be an azimuth estimated based on TDOA, $a_k(t)$ be an azimuth estimated from Kalman filter and s(t) be the value of cross-correlation at time t. We calculate the final azimuth $a_f(t)$ like this.

$$a_f(t) = s(t)a_o(t) + (1-s(t))a_k(t)$$  \hspace{1cm} (8)

If cross-correlation value is large, the equation weights the final azimuth with the azimuth estimated based on TDOA. If cross-correlation value is small, the equation weights the final azimuth with the azimuth obtained from Kalman filter.

### 3. Experiment

#### 3.1 Environment

In this paper, we make an experiment with iCat made by Phillips International, Inc[11]. Two microphones are detached in each foot. The distance between two microphones is 22 cm. The sampling rate of sound signal is 4 kHz, 8 kHz and 16 kHz in each experiment. The threshold of energy is 0.0025 and the window size of the cross-correlation is 100 frames. We use the footsteps sound as an impulse sound. We use the Kevin Murphy’s Kalman filter toolbox[15].

#### 3.2 Experiment result

We make two experiments. First, we test the accuracy of azimuth result. Second, we test how the azimuth result is natural. In the first experiment, we walk along the predefined nine points in front of robot. Table 1 shows the result.

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Method 1</th>
<th>Method 2</th>
<th>Method 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 kHz</td>
<td>23.20</td>
<td>21.92</td>
<td>21.97</td>
</tr>
<tr>
<td>8 kHz</td>
<td>10.21</td>
<td>10.20</td>
<td>9.45</td>
</tr>
</tbody>
</table>

Table 1 The RMSE (Root mean squared error) of the estimated azimuth result. Method 1 is based on only TDOA. Method 2 is using Kalman filter based on TDOA. Method 3 is weighted Kalman filter using cross-correlation value based on TDOA.
In the table 1, method 1 is based on only TDOA. Method 2 is using Kalman filter based on TDOA. Method 3 is weighted Kalman filter using cross-correlation value based on TDOA. We can see that the input sound signal with a high sampling rate gives the more accurate azimuth. And the proposed method 3 is more accurate than other approaches.

In the second experiment, we test that how the estimated result is natural. We walk repeatedly in front of the robot and see the trace result of azimuth. The result is shown in Fig. 2.

![Fig. 2](image)

In the Fig. 2, we can obtain more natural azimuth, when we use a sound signal with higher sampling rate. And in the middle of Fig. 2, the method 2 and 3 can alleviate the result calculated from mis-obtained sound signal using only TDOA. Method 3 gives more natural result than other methods.

4. CONCLUSION

We proposed the method which calculates the azimuth of the sound source and tracks it on the robot environment with two microphones. We calculate the azimuth based on TDOA and give chase to the azimuth of the sound source. Finally we update the azimuth of sound source using the cross-correlation value. In the experiments, we showed the proposed method is more accurate and natural than other methods.

REFERENCES


