Sound Source Localisation Assisted Active Security Camera

Wijekoon M.P.A., Wjesinghe W.T.P., Upendra G.A.A., Vitharana U.B.
Supervised by. Dr. Thrishantha Nanayakkara
Department of Electrical Engineering, University of Moratuwa,

Abstract **-The main objective of this project was to utilize Sound Source Localization technique for constructing an Active Security Camera.**

I. INTRODUCTION

Most Security cameras installed in premises are not active. They are only supposed to record the activities, of certain directions continuously. Therefore a security camera, which is active and has the ability of steer itself towards a suspicious noise will certainly has an advantage when competing with common security cameras in the market. The ultimate objective therefore was to build and market a Low Cost Active Security Camera, which can be used in industrial and domestic premises.

II. PROJECT OVERVIEW AND IMPLEMENTATION

The system uses SSL technique to estimate the location of an acoustic source, which in this system would be a sound originating due to any suspicious activity. Once estimated, a camera is steered to target the estimated position. A Passive Infrared (PIR) sensor is employed to capture the Infrared radiation at the estimated position and is used to verify the existence of a PIR source at the estimated location. If such a body is detected, which will be the case if a person is present at that location, a security breach is assumed and, control signals are activated, which can be used to take appropriate emergency security measures.

In addition to this mode of operation, it is designed to track a known sound source. A specifically designed acoustic source is used, which emits signals in the range 12 KHz to 20 KHz. In this mode camera will track this source in a limited area and no use of the PIR sensor module is made.

Possible applications for this mode are tracking of a speaker, robot, or any other object, which requires to be tracked in a specific area.

To realize the localization several steps are been taken. The first step is the sensing of acoustic source through four microphones. This data is acquired and processed using algorithms implemented in software.

Once acoustic position is estimated, this data is used to drive two stepper motors, which control the camera angle. Communication between PC and this part of the hardware is done through serial port. If operated as a security camera, in addition to motors, PIR sensor module is also controlled through the same port.

III. TDOA BASED LOCATORS

The conventional TDOA method can be separated into two steps.

Step 1:- Time delay estimation (TDE) is calculated for each pair of microphones. This step is performed either in time domain as the cross correlation between two signals or using generalized cross-correlation (GCC) method. The GCC method is discussed further in the report.

Step 2:- The estimated time-delays for pairs of microphones can be used for getting the location of the sound source.

IV. DELAY ESTIMATION

The well-recognized Phase Transform algorithm is used in the estimation of delays.

V. MODULE FOR THE ESTIMATION OF SOUND SOURCE POSITION

Three different mutually exclusive techniques are proposed and applied in the implementation of this section.

One of them, which required minimum processing, is to develop a closed form expression for source position coordinates in terms of time delays. Major disadvantage of this technique is estimated position is highly dependant upon the accuracy of the time delays. A slight deviation of these delays from its exact value will result in large errors in the position estimation. In extreme cases, equations become even indeterminist.

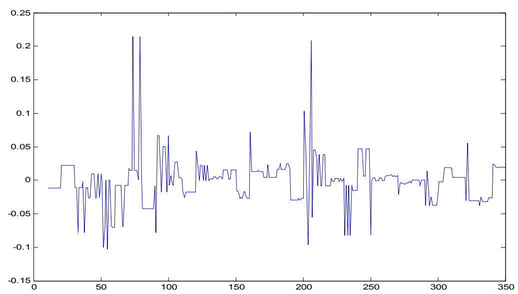
Second technique is to use a lookup table to determine the locations. This requires a previous data set, which comprises estimated delays with the actual location of the source corresponding to the delays. The table lookup is a linear search for Least Square Error (LSE). This requires more processing than first method, but it always has a solution and is not very sensitive to slight variations in delays. Once implemented this method worked well, although to cover significant area, larger number of actual data sets should be available.

A third technique is based on an Artificial Neural Network (ANN), which maps input delays to location.

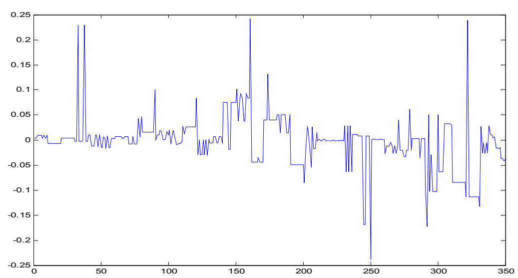
A. Neural network based implementation

A suitable network is designed, which gives a Mean Squared Error (MSE) of 1.1 for a training set of size 350.

Following two graphs show the error plot for this training set. Maximum error is about 20% of the output. Table-1 shows percentage of training data, which have a percentage error less than the given threshold value.



Error Plot for Tilt Angle



Error Plot for Pan Angle

Table 1 Error variation

% Error threshold	% of training set with error less than the threshold (Pan angle)	% of training set with error less than the threshold (tilt angle)
.1	53	51
1	71	71
5	97	91
10	99	98

From the table it is clear that, although, maximum error can go as high as 20 %, number of training data having that much of deviation is very small. If an error of 5% is acceptable, about 97% for pan angle and 91% for the tilt angle calculations will lie inside this limit. This is acceptable for this system and work is carried out to implement the network in Visual C++. Table-2 shows the structure of the resulting neural network.

Table 2 Structure of the network

Layer	Number of Nodes	Transfer function of nodes
Input	3	Log sigmoid
Hidden layer # 1	15	Hyperbolic tangent sigmoid
Hidden Layer # 2	20	Log sigmoid
Output	2	Linear

Processing power required for this method is low compared to the table lookup. In addition, a properly designed network structure will realize the actual relationship between delays and position thereby giving accurate results for new data, which were not in the training set. However this is still not tested with the designed neural network.

Implementation of this network is currently being carried out using Visual C++. An object-oriented approach is taken in the implementation, where different classes are defined to encapsulate the functionality of Network, Layer and Node. These objects can be combined to build the network incrementally.

B. Derivation of closed form expression

Many researches have been carried out using array of microphones to acquire the position coordinates. Given the position coordinates of two microphones and the TDOA between the microphones, an equation can be formulated giving the relationship between the two parameters (TDOA & Sensor Positions) and the acoustic source position. This equation presents one half of a hyperboloid of two sheets. The hyperboloids axis of symmetry is the line joining the two microphones. The source point can be on any point of the hyperboloids plane. Hence to get 3D location there should be at least three intersecting hyperboloid planes. Therefore the minimum number of microphones needed is three and, of course increasing the number of microphones will increase the number of intersections at the source point. With this method an iteration method must be employed to minimize the error of false point localization due to reverberations and multiple intersection locations cause by the hyperboloid planes.

C. Table lookup based implementation

In this implementation a table is constructed with fields as follows.

Delay 1	Delay 2	Delay 3	Tilt Angle	Pan Angle

Delay 1, Delay 2, and Delay 3 are the three delay values estimated from the delay estimation module. Tilt and Pan Angle are the angles of the sound source location based on a fixed spherical coordinate system

These two angles can directly be used in the mechanism control module, which uses them to rotate two stepper motors in horizontal plane and vertical plane.

Crucial factor for the successful operation is to construct a comprehensive table to cover a wide area. The procedure (call this procedure "training") involves operating the sound source at a known tilt and pan angle and running the system to compute the three delays. These delays along with the tilt and pan angles are saved in the table. About ten readings are taken for each position to allow slight variations in the delay estimation to be included in the table. This table is saved permanently in the disk and loaded during real operation.

Criteria behind the selection of an entry for a given time delays is the Least Squared Error(LSE), which is defined as,

$$LSE = ((dt_1)_c-(dt_1)_s)^2 + ((dt_1)_c-(dt_1)_s)^2 + ((dt_1)_c-(dt_1)_s)^2$$

where $(dt_i)_c$ for i=1,2,3 denotes currently estimated delays and $(dt_i)_s$ for i=1,2,3 denotes saved delays corresponding to the current entry in the table.

LSE is calculated for all values of the table and the entry corresponding to the minimum is taken as the estimation. The other two fields of this entry contain the position coordinates of the source as tilt and pan angles, which are used in subsequent modules.

VI. MOTOR MECHANISM

The final out put of our project is to accurately direct the camera to the sound source according to the calculated data from the computer. This final step is achieved by developing a mechanics, which has the movable capabilities in the 360-degree solid space. The basic design consists of three parts namely are,

- The base of the mechanism
- The intermediate part
- The tray

The above mentioned parts are carefully designed to make the structure lighter and strong while having its smooth maneuverable capabilities to the (tilting and panning features) to the mechanism. The parts of the mechanism are coupled together in a way that they can revolve around vertical and horizontal axes that gives the mechanism its 360 solid angle flexibility. The next few paragraphs explain the functionality of the mechanism. The base is holding the weight of the other parts and it can be mounted horizontally (in standing or hanging positions) or vertically on a wall. This part is equipped with a stepper motor, which drives a vertical shaft that gives the panning feature to the mechanism.

The vertical shaft is then coupled to a gear wheel and it was directly coupling to the motor gear wheel. But because of some mismatching of the gear-teeth it produces a noise and this fault was corrected by introducing a belt between the two gear wheels.

This vertical shaft is then coupled to the intermediate part that can revolve 360 degrees around the axis. This part holds a smaller stepper motor and the tray driven by this motor. The weights of the components are distributed symmetrically to minimize the load variations to the motor while it's operation. The tray is balanced no the intermediate part of the mechanism in a stable equilibrium position that gives the smooth titling feature to the tray. The revolving axis of the tray was also placed in such a manner that the weight of the camera and other gadgets inside the tray will not make any considerable variation to the motor load while its 180-degree rotation.

There are position sensors placed on the base and on the intermediate part as well as in the tray to get a reference point for the panning and the tilting movements. The next important thing was to supply power and controlling and input signals to the motors and to the camera and return the sensor outputs to the computer (serial port interface). To make this task easier a terminal bar was fixed inside the intermediate part where all the power and controlling signals from the apparatus are connected to it and a bundle cable is taken out from that to the controlling circuits. The base has a similar arrangement to pass the input and output signals between the apparatus and circuits.

Finally the camera mechanism was mounted on a wooden cubical and place all the circuits (that drives the motors and other controlling circuits) underneath the mechanism, where a bundled cable connected between the two carries all the signals back and forth.

D. Camera Position Sensors

Although stepper motors are used to drive the camera, at least one known position should be available for proper positioning. In order to identify this position (Call it Reset position) two sensors are used. Reason for having two sensors is that, camera move both in horizontal and vertical plane. Two Infrared emitter detector pairs are used.

E. Passive Infrared (PIR) sensor

A sound may be generated by a person or without any intervention by a person. This latter situation arises most of the time due to animal movements, wind etc. If these two situations are not differentiated properly, system will unnecessarily disrupt the user for noises originating from un-harmful sources. In order to achieve this, a Passive Infrared (PIR) sensor is used. Objects that generate heat also generate infrared radiation including animals and the human body. PIR sensors provide significant output voltage variation (when used with an appropriate circuit) when exposed to such a body. This fact can be used in differentiation, as sound generated due to wind or some other similar source will not emit any Infra red at the source location. Small animals will emit some radiation, but it is much less than a person. Therefore it is easy to differentiate a person from some other source by analyzing the out put of the PIR sensor. Only additional requirement for this to function is that the PIR sensor should be directed toward the location of the source, which is achieved by coupling it to the camera itself.

VII. SYSTEM OPERATION

As briefly mentioned in a previous section, system supports two modes of operation.

- Tracking of a known sound source
- As a security camera

Above two operating modes are described in following sections.

F. *Tracking mode*

In tracking mode a gradually alternating sine wave in the frequency range 12 KHz to 20 KHz is used. This type of sine wave with changing frequency is used to avoid periodicity. If periodicity is present in signals, which are used in, cross correlation calculation, the result itself will be periodic with a time period equal to the time period of input signals. This will result in many local maximums, which complicate and even prohibit the determination of time delay accurately.

G. *Security camera mode*

As a security camera, no control over the sound source can be exercised. Instead it is an arbitrary sound, which may be periodic or non-periodic, and may have arbitrary power spectrum. However periodic signals are unlikely to occur in this application area, hence no major consideration is given to avoid the effect of such signals.

In this mode, both hardware and software filters are disabled. Camera is steered to target the detected source. Just before the target position is reached, external Analog to Digital Converter is started to sample the output of the PIR sensor. This output is processed by the microcontroller in order to verify the existence of a PIR radiation body at the source location. More details about this verification process are given in section 4.8.2. Result of this analysis is then send to the PC. Based on this result, relevant modules are activated to take appropriate actions.

There are two outcomes for PIR sensor analysis.

A PIR source is detected at the target location:

In that case, post processing module is signaled about the possible security breach. This module can be programmed to operate a video recorder, activate an alarm, send a message to a mobile phone, call a preconfigured phone number, etc.

No PIR source is detected at the target location:

In this case, it is assumed that no security breach is occurred. No further actions are required.

This type of verification process is introduced to reduce unwanted activation of emergency security systems even for unharmfull sounds sources. Detection sensitivity is configurable by the user. If configured properly, it can even be used to detect a person while avoiding detection of animals like cats and dogs.

Unlike in tracking mode, this mode supports default movements for the camera. In this way user can change the movement of the camera during normal operation; i.e. before detecting any suspicious noise. It can be stationary, in which case camera stays at its reset position, rotating by some angle at a time or any other type of behavior the user defines. User can even control the camera using the mouse until the required scene is captured. These are the features provided by most generic security camera systems. In this system, that type of functionality accounts only for a small fraction of the whole system.

VIII. RESULTS

System is first tested in tracking mode and using table lookup for position estimation. Very good results are obtained even under low SNR. During this operation, most of normal noises were present in the environment. This included Air conditioner noise, fans and people talking.

A resolution of approximately 30 cm at horizontal plane is achieved. Camera position is updated at an approximate frequency of 1 Hz, with a 550MHz Intel Pentium III Processor.

Severe reverberation present in the room in which the system is located has resulted in reduced performance for the other mode of operation relative to the tracking mode. With higher frequency signals used in the tracking mode, reverberation is low and doesn’t significantly affected delay estimations. Completion of functionality in this mode is hence deferred as a possible future extension. This may require some alterations to the Phase Transform Algorithm and measures to reduce the reverberation in the environment.

ACKNOWLEDGMENT

We would like to take this opportunity for acknowledging personals who helped us in completing our final year project.

Our deepest gratitude goes to our project supervisor Dr. Thrishantha Nanayakkara who provided us with abundant of support throughout the project. Dr. Thrishantha facilitated us, in both intellectually and materially throughout the project.

We would like to acknowledge Head- Department of Electrical Engineering Dr. Ranjith Perera and Prof. J.R. Lucas (Former Head of Department) for granting us the use of Power Electronics laboratory for the project work.

REFERENCES

[01] *Robust Sound Source Localization Using a Microphone Array on a Mobile Robot*
Jean-Marc Valin, Franc¸ois Michaud, Jean
LABORIUS - Research Laboratory on Mobile Robotics and Intelligent Systems
Department of Electrical Engineering and Computer Engineering
University of Sherbrooke
Quebec, Canada

[02] *CMSC 660 Project Solutions Optimization methods for Sound Source Localization using Microphone arrays*
Vikas C. Raykar
December 19, 2002

[03] *A DSP Implementation of Source Location Using Microphone Arrays*
Daniel V. Rabikin, Richard J. Renimeron, Arthur Dahl, Joseph C. French, James L. Flangan
Ctr. For Computer Aids for Industrial Productivity

[04] *Sound Source Localization in Reverberant Environments using an Outlier*
Elimination Algorithm
Ea-Ee Jan yand James Flanagan
CAIP Center, Rutgers University, Piscataway, New Jersey 08855

[05] *New Direct Approaches to Robust Sound Source Localization*
Yong Rui and Dinei Florencio
Microsoft Research
One Microsoft Way, Redmond, WA 98052

[06] *Talker Tracking Display on Autonomous Mobile Robot With a Moving Microphone Array*
Takanobu Nishiura, Masaya Nakamura, Akinobu Lee, Hiroshi Saruwatari,
Faculty of Systems Engineering,
Wakayama University
930 Sakaedani, Wakayama, 640-8510 Japan
ATR Spoken Language Translation Research Labs.
Graduate School of Information Science,
Nara Institute of Science and Technology

[07] *16F877A Data Sheet*
.
[08] *MATLAB Documentation*