EIGHTS INTERNATIONAL CONFERENCE ON APPLICATIONS
OF ADVANCED TECHNOLOGIES IN TRANSPORTATION
26-28 MAY, 2004 BEIJING, CHINA

Automatic Traffic Camera Steering for Incident Imaging Using an Array Processing Technique

Abedin Vahedian
Khorasan Research Center for Technology Development
Mashad, Iran
vahedian@kstp.ir

Abstract
When an incident occurs at the cross sections where traffic cameras are installed, it is expected that the operator in traffic control center attends the event to steer the camera to monitor the incident and/or start recording the scene for further investigations. This is one fairly easy task provided that the operator at the traffic control center is present and able to track the incident. During nights and other occasions, however, where the operator is not able to fulfill the task, there might be a need to be able to have a local control on the traffic cameras to apply proper pan/tilt and zoom commands to steer the camera to focus at the scene of incident. A sensor-array based technique presented in this paper uses the acquired signal from a tiny ultrasonic whistle activated by the police officer attending the incident. It uses the signal to estimate the position of ultrasound source, i.e. the police officer on the street and then steer the camera toward it.
Key words: source location, direction of arrival (DOA), beam-forming, array processingg

1. Introduction

Traffic cameras not only facilitate monitoring of traffic flow on the highways and cross sections but also are a very unique tool to acquire live video information upon occurrence of an incident. In such situations the operator in the traffic control center should take a series of necessary steps, namely steering the closest camera to the scene and focus on the event. At times, however, when operator is not available an automatic steering function devised on the camera might be of highly desirable for the policemen attending the scene. In this scenario, a technique to enable the camera to receive pan/tilt and zoom functions from the scene to steer toward it in the field of monitoring could be of great interest. To do this, it becomes necessary to identify and track a specific object in the field of interest in real-time fashion. This often becomes a major requirement in applications such as machine vision, security and monitoring.

Much research effort has been addressed at detection and direction of arrival estimation problems. Most has attempted to use properties of the video image acquired from scene to identify the location of a specific object, such as a collision, which is then used to enable the monitoring system to track or apply further and complementary processing [1, 2]. These techniques tend to be computationally very complex and make assumptions about the content of the image. It turns out that to accurately locate an object using the video information alone is a very difficult problem. On the other hand, it is possible to use an external source of information, namely the signal information emanating from a source in the scene to identify its location and then automatically steer the camera toward the source with proper zoom.

The ultrasound-assisted technique presented in this work is based on the idea of beam forming. Hence, by combining the acoustic source direction information, it is possible to achieve a significant increase in utilization of traffic video cameras without a considerable expense. The approach implemented thus far, has been to use an array of sensors to find the direction of arrival of ultrasonic source using DOA estimation schemes.

In the next section, the concept of using a sensor array in this particular application is explained. The implemented design to carry out the experiments is then presented. Performance evaluation in using the acoustic information is then discussed. Finally, the paper is concluded with the future work program toward the complete implementation of the proposed system.

2. Identifying Ultrasound Source Using an Array of Sensors

In an acoustic DSP system, the directivity of a sensor array can be used to pick up acoustic signal emitted by a distant source while suppressing noise and reverberation arriving from other directions. There are many direction finding techniques and all these techniques exploit the fact that the time taken by a signal emitted by a source to reach different sensor is different due to the spatial spread of the sensors [3, 4, 5]. The position of an acoustic source may be located by considering different points on a spiral starting at the outer boundary of the area under consideration and measuring the delay in arriving from each point using the optimal delay estimation methods. Fig.1 shows a general set up.

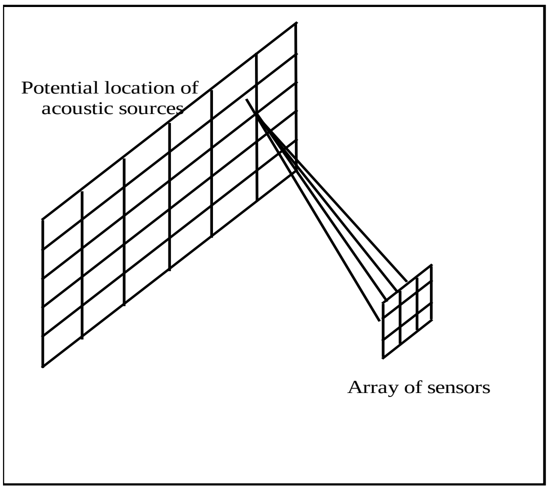

Fig.1 Array of sensors used to cover the area of interest.

The scheme uses the time difference of arrival (TDOA) estimates between the source and a spaced array of sensors forming a planner array. This difference is easily converted to the range difference since sound waves propagate at a relatively constant speed, at least inside a larger room. Therefore, each TDOA measurement between two sensors determines that the position of sound source must lie on a hyperboloid with a constant range difference between the two sensors. Fig.2 depicts the solution where two hyperboloids are formed from TDOA measurements at three fixed sensors to provide an intersection point that locates the position of source. The equation of this hyperboloid is given by:

$$R_{i,j} = \sqrt{(X_i - x)^2 + (Y_i - y)^2 + (Z_i - z)^2} - \sqrt{(X_j - x)^2 + (Y_j - y)^2 + (Z_j - z)^2} \qquad (1)$$

with $R_{i,j}$ being the range difference between the sensors *i* and *j*. The coordinates (X_i, Y_i, Z_i) and (X_j, Y_j, Z_j) represent the two fixed sensors and make up the unknown coordinate of the sound source position.

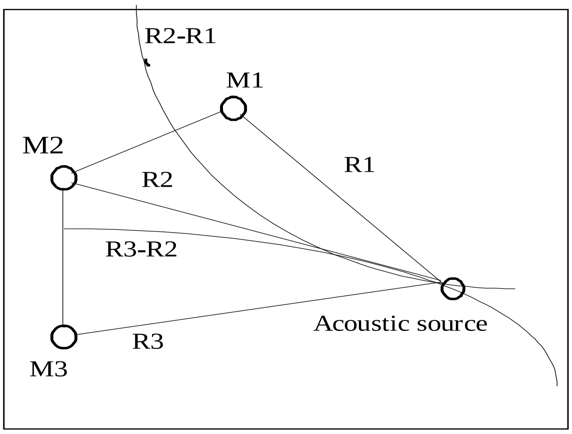

Fig.2 Hyperbolic position location with 3 sensors for a 2-D solution.

In the general case, in which the 3-D position of the source is required, at least four independent measurements need to be made.

The technique is implemented in two stages: First, estimates for TDOA values are computed from sound signals. Then estimated TDOA values are processed to determine a direction estimate. These steps are now explained:

1-*Computing TDOA Estimates:* To evaluate the hyperbolic range equations, we need to obtain the estimate of the range differences, or equivalently the TDOA.

A narrow-band signal emanating from a distance source and monitored in the presence of noise at two spatially separated sensors can be mathematically modelled as:

$$\begin{aligned} x_1(t) &= s_1(t) + n_1(t) \\ x_2(t) &= \alpha s_1(t + D) + n_2(t) \end{aligned} \qquad (2)$$

where the signal and noise are assumed to be real, jointly stationary random processes. Signal *s1*(t) is also assumed to be uncorrelated with the noise signals.

One common method of determining the time delay (*D*) and hence, the arrival angle relative to the sensor axis is to compute the cross-correlation function [6]

$$Rx_1x_2(\tau) = E[x_1(t)x_2(t-\tau)], \qquad (3)$$

where E denotes expectation. The argument that maximises (3) provides an estimate of delay. This implies that in implementation, the A/D and the data logger must share a precise time-base with the reference signal but does not impose any requirement on the signal transmitted by the source. Because of the finite length of recorded samples, however, $Rx_1x_2(\tau)$ can only be estimated. In order to improve the accuracy of the delay estimate, it is necessary to prefilter x_1 and x_2 prior to the cross-correlation process. An estimated cross-spectral density function can also be computed in the frequency domain, and then the estimated cross-correlation function is obtained via an inverse Fourier transform. The later approach has been adopted in this study as the frequency domain processing lends itself well to filtering of the signal prior to computation of the cross-correlation function.

The Phase Transform (PHAT) processor has been used for this purpose:

$$\hat{R}x_1x_2(\tau) = \int_{-\infty}^{\infty} \frac{\hat{G}x_1x_2(f)}{|Gx_1x_2(f)|} e^{j2\pi f\tau} df. \qquad (4)$$

where the expression $Gx_1x_2(f)$ denotes the cross power spectrum of each pair of signals. For the model in (2) with uncorrelated noise (i.e. $Gn_1n_2(f) = 0$),

$$|Gx_1x_2(f)| = \alpha Gs_1s_2(f). \qquad (5)$$

and for a relatively long time monitoring, when
$\hat{G}x_1x_2(f) = Gx_1x_2(f)$, then

$$\frac{\hat{G}x_1x_2(f)}{|Gx_1x_2(f)|} = e^{j\theta(f)} = e^{j2\pi fD} \qquad (6)$$

and the cross-correlation estimation approaches a delta function. In practice, performing the above delay estimator on a pair of relatively long sound samples (compared to the sample delay expected) leads to a peak in the cross-correlation function.

2-Direction Estimation by solving the Hyperbolic Equations: Once a reliable estimate for TDOA values has been computed, it is substituted into the hyperbolic equations, which are then solved for the Cartesian coordinate of the acoustic source, with respect to a set of objects (i.e. the coordination of sensors, or one reference sensor assumed on the (0, 0, 0) location). At this stage z is assumed to be $1m$. It however does not affect the generality of procedure and is just for sake of simplicity. Actual depth i.e., z coordinate, is variable is and solely dependent to zoom factor.

Solutions for hyperbolic and related position fixes have attracted attention for applications in mobile stations and navigation systems [7,8]. In traffic monitoring application, however, there is not a great challenge for a higher resolution within a whole of scene range of $\pm 50m$ broadside; neither is of importance the processing time of the algorithm. Therefore, assumptions exist on the dimensions of the area of interest for monitoring which allow us to reduce the complexity of the computations to a linear level, as the hyperbolic equations are non linear and their solution is nontrivial, particularly when range estimates may be inconsistent due to the noise or multipath arrival or echo.

The direction estimate from the above two step procedure is now converted to the proper commands for the camera pan/tilt controller as vertical/horizontal directions according to its current position. where one single angle within the field of view is determined as being the direction of the acoustic source. For zooming, strengths of the received signal is taken as a measure for distance and so considering a squared relationship between the signal strength and distance, the zoom is set to the peak value of the signal, the weaker the received signal the bigger the zoom factor.

3. Implementation of the Experiment

A rectangular array of four ultrasound receivers spaced *20* cm from each other has been used in this work. The camera is located in the center of the sensor plane such that its focal plane and the sensor plane are the same as shown on Fig.3.

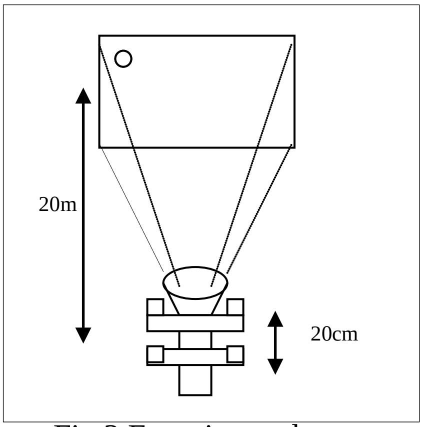

Fig.3 Experimental set-up

Four ultrasound channels have then been transferred to the computer through a *4*-channel simultaneous A/D data acquisition card. A *4*000 sample section of each analogue channel was used in the TDOA estimation procedure. At a sampling rate of *60ks/s*, a time delay equivalent to one sample then represents a range difference of approximately 0.5*cm* between two sensors, which in turn represents a range difference of about *50cm* along the scene where an ultrasound transmitter playing a 30-*khz* tone was placed in the scene as the acoustic source at a distance of *20m* from sensor array.

4. Results and Discussions

In the acoustic signal processing section, using PHAT provides reliable TDOA estimation results. When the source was displaced in the scene or the camera made panning over the scene, a displacement of ±25cm was detected. The projection task performed in software according to the known geometry and the coordination reference also made fairly reliable direction identification. A video sequence recorded on tape will accompany this paper that clearly demonstrates the impact of the direction of arrival estimation algorithm on the self steering capability of camera.

Further work to be carried out as part of thee proposed technique includes:

* techniques to analyse the aacoustic data so as to achieve the best minimum resolution in terms of detectable range in a scene of interest.
* cancelling echo and reverberation and resolving multipath arrivals of signal using efficient but low cost beam-forming techniques.
* increasing the sensible range in both the vertical and horizontal directions considering large variations in zooming factor.

5. Conclusions

A new idea has been demonstrated to exploit the relationship between sound and camera control in traffic video applications. It has been shown that the performance of a traffic surveillance system task can be improved with little increase in the overall complexity. The continuation of this work can open up a new dimension in using every possible correlation between external non video factors and image video analysis in traffic video based applications.

6. References

[1] R. Mech and P. Gerken, “Automatic Segmentation of Moving Objects”, Doc. ISO/IEC JTC1/SC29/WG11 MPEG97/1949, Bristol, UK, April 1997.

[2] B. Falcidieno and T. L. Kunii (Eds), “Modeling in Computer Graphics”, Springer, Berlin, 1993.

[3] B. Ottersten and T. Kailath, "Direction of Arrival Estimation for Wide band signals ...", IEEE Trans, ASSAP, Vol. 38, No. 2, Feb. 1990, pp. 317-327

[4] S. R. de Graaf and D. H. Johnson, "Capability of Array Processing Algorithms to Estiamte Source Bearing", IEEE Trans. ASSAP, Vol. 33, No.6, Dec. 1985, pp. 1368-1379.

[5] Barry D. Van Veen and Kevin M. Buckley, "Beamforming: A versatile Approach to Spatial Filtering", IEEE ASSP Magazine April 1988.

[6] C. H. Knapp "The Generalized Correlation Method for Estimation of Time Delay", IEEE Trans. ASSP, Vol. ASSP-24, No.4, August 1976.

[7] B. T. Fang "Simple Solutions for Hyperbolic and Related Position Fixes", IEEE Trans. Aero. and Elec. Sys., Vol. 26, No. 5, Sep. 1990.

[8] T. S. Rappaport et al, "Position Location Using Wireless Communications on Highways of the Future", IEEE communications Magazine. October 1996.

[9] A. Vahedian, et al, “Improving Videophone Subjective Quality Using Audio Information”, International Journal of Imaging Systems and Technology, Vol 10, 86-95 (1999).