# Robust Binocular Tracking with an Auditory Perception System

T. I. James Tsay and Yuan-Yu Li

*Department of Mechanical Engineering, National Cheng Kung University*
*Tainan, Taiwan 701, R.O.C*
*tijtsay@mail.ncku.edu.tw*

***Abstract***: This paper uses an auditory perception system to assist visual tracking. Using auditory perception system to track the source of the sound from a target and then moving the robotic head to face the target help the CCD cameras of the head locate the target's image, ensuring that the visual tracking works. Combining the visual tracking system with an auditory perception system increases the overall efficiency of the visual tracking system. In auditory perception, an equilateral triangular array of microphones is used to receive sound signals, and the direction to the sound source is determined using the cross-spectrum method. Digital signal filters and a cepstrum lifter can be used to recognize the digital signal to determine whether the sound matches the target's sound model. Then, the target's sound source is recognized and its direction is determined. In image tracking, three-step hierarchal search method is replaced with the edge-blobs pattern matching method proposed herein to increase the tracking speed. It is combined with edge-blobs contour matching and YCbCr space image recognition to improve the recognition of the target image and to filter out background noise. The target is then tracked regardless it is out of shape or passes through a work space with complex background. A robust binocular tracking system with an auditory perception system is developed to enhance target tracking. One experiment is conducted to verify the effectiveness of functions for the developed system.

***Keywords***: auditory perception, binocular, cepstrum lifter, edge-blobs pattern matching, edge-blobs contour matching.

## I. INTRODUCTION

With advances in robotic vision, visual target tracking remains constrained by difficulties in tracking invisible targets: designing an image tracking algorithm to search for a target that is not in the image plane is difficult. If, when the tracking system starts, the target is not in the image plane, or it is covered by objects in the working space, "target disappearance" may occur and much time will be spent searching for the target in the work space, causing the visual tracking to become unstable.

Some recent studies about visual tracking, such as those of Chen [1] and Wang [2], have combined pattern matching and contour matching to increase the robustness of visual tracking systems. In 1995, Yanagisawa et al. [3] constructed an equilateral triangle microphone array to track a whistle. In 2002, Chen [4] realized such a sound tracking system by designing IC circuits and using digital signal processors. He placed this sound source tracking system on an autonomous guided vehicle to enable it to track a whistle.

While combining auditory perception with visual tracking, Hu et al. [5] developed a self-calibrated speaker tracking system based on both audio and video information. They set an array of equilateral triangle microphones on a pan-tile active camera to track a speaker. The Japanese company, NEC, researched and designed a small, intelligent robot-R100 with a height of 44cm and a weight of 7.9Kg. It has two cameras, three microphones, many force sensors and ultrasonic sensors. NEC began to plan the R100 in 1997. The robot-R100 combines pronunciation recognition, image recognition and an electromechanical system to track and recognize specific people. R100 can understand hundred Japanese characters. This robot responds appropriately, enabling simple interactions with humans.

In this paper, a simple but fast image-tracking algorithm – edge-blobs pattern and contour matching is proposed for visual tracking. A color space target model is constructed and color space patterns and contours matched to filter out background noise. In auditory perception, an equilateral triangular array of microphones is used to track the sound source and perform simple voice recognition. The sound source-tracking algorithm separates the two-dimensional-pan space into 12 directions. The goal of sound source tracking is to control the robotic head to face the target to track the image. In summary, a robust binocular tracking system with auditory perception is developed to enhance target tracking. This control architecture can be used not only in a security system but also in home robots to improve their interaction with humans.

## II. IMAGE TRACKING

The algorithm combines edge-blobs pattern with edge-blobs contour, such that it combines the inner information of the target with the outer information. This algorithm incorporates the YCbCr color space into this algorithm, as shown in Fig. 1. This cue method outperforms pattern matching alone, contour matching or gray level space image because it finds not only the degree of matching between the inner information of the target but also finds them between the outer information, and the YCbCr space is used to filter the objects whose colors differ from that of the target. This improves performance when the background is complex, and reduces the processing time because edge-blobs pattern matching and edge-blobs contour matching have few blobs than gray level matching. Accordingly, it is a fast, high-performing image tracking framework.

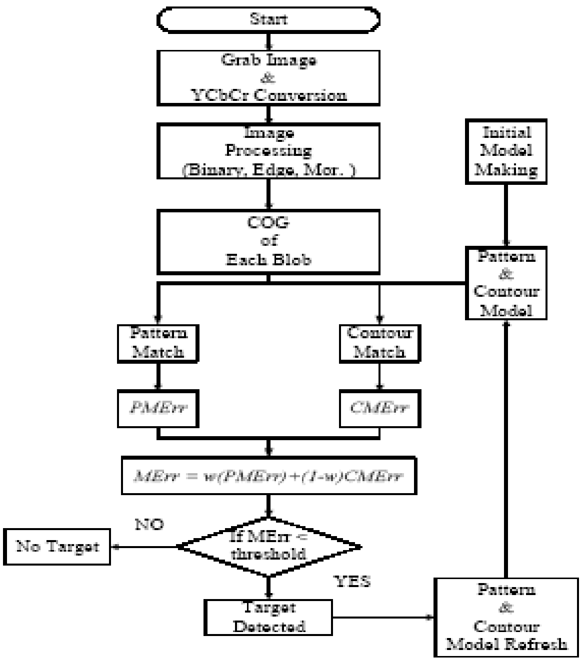


Fig. 1 Image tracking algorithm

## III. SOUND SOURCE TRACKING

A sound sensor, or "microphone", can detect only the magnitude of sound intensity. If two sensors are present, sound phase lead-lag in the direction of the separation of the microphones can be determined. More sound sensors allow more directions to be defined. Accordingly, a sensor array is made by arranging microphones in a specific geometrical pattern, such as a circle or an equilateral triangle. Fig. 2 displays an equilateral triangle with three microphones used herein, to elucidate the efficiency of signal sampling and the angular resolution of the sound source.

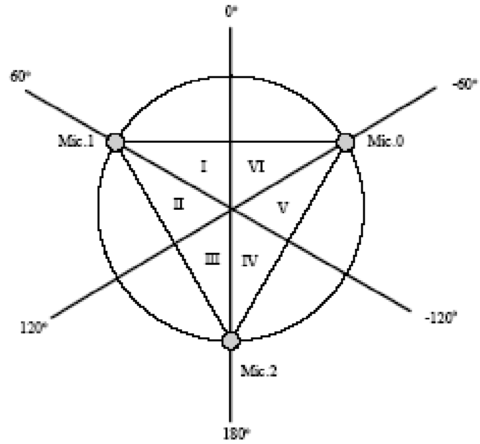


Fig. 2 Array of Microphones

Sound source tracking involves the following four steps.

Step1: Use an array of microphones to obtain sound signals from the workspace.

Step2: Convert the analog signals to digital signals and filter them through digital signal filters.

Step3: Perform sound source recognition using the cepstrum method to check whether the signals are emitted from the target to be tracked.

Step4: Perform the cross-correlation for each of the signals from each microphone and calculate the signals' phase lead-lag degree to determine the direction to the sound source. Sound source tracking is thus completed.

## IV. BINOCULAR TRACKING WITH AUDITORY PERCEPTION

### 4.1. Hardware and Control Architecture

The target-tracking system is a five-axes robotic binocular head with an equilateral triangular array of microphones, which has sides of 28.5cm. The overall target-tracking system displayed in Fig. 3 combines the sound source tracking control system with the visual tracking control system. PC1, as presented in Fig. 3, tracks the sound source and obtains the motors' encoder signals and the image tracking commands; finally, it sends suitable motor commands to keep the target in the centers of image planes.

When the tracking system starts, the first step of the tracking algorithm searches the image target in both of the CCD image planes. If the image target is present,

The Thirteenth International Symposium on Artificial Life and Robotics 2008(AROB 13th '08),
B-Con Plaza, Beppu, Oita, Japan, January 31-February 2, 2008

then the second to fifth axes of the motors maintain the image target in the center of the image plane; otherwise, the sound source is tracked until the target appears in one or both of the CCD image planes. Fig. 4 shows the overall tracking algorithm.

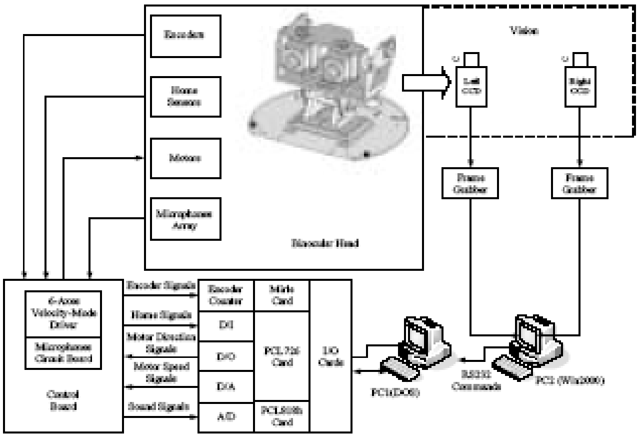

Fig. 3 The overall target-tracking system

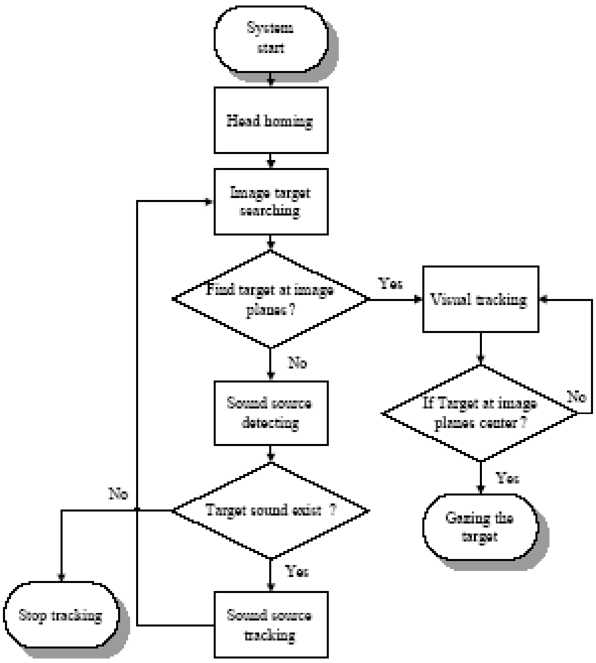


Fig. 4 Block diagram of overall tracking system

### 4.2. Experiments

The experiment concerns binocular tracking with the auditory perception system under complicated and semi-covered circumstances. Fig. 5 depicts the experimental setup. A cover plate is placed to cover the left-half path of the target. The goal of this experiment is to test the robustness of the overall tracking system, as the speed and the direction of the covered target changes. The target is covered from 0 cm to 20 cm along the sliding rail, so a tracking error of around 10cm exists from time 0sec to 15sec. From 15sec to 20sec, the target accelerates. The experimental results in Figs. 6 to 8 display the robotic binocular head tracking performance.

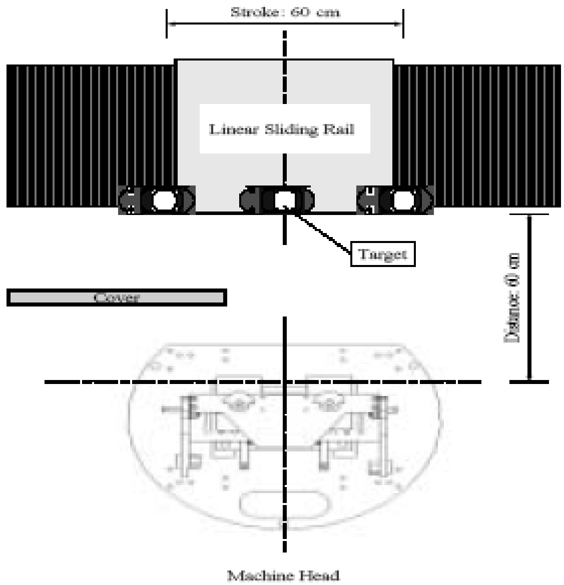


Fig. 5 Experimental setup

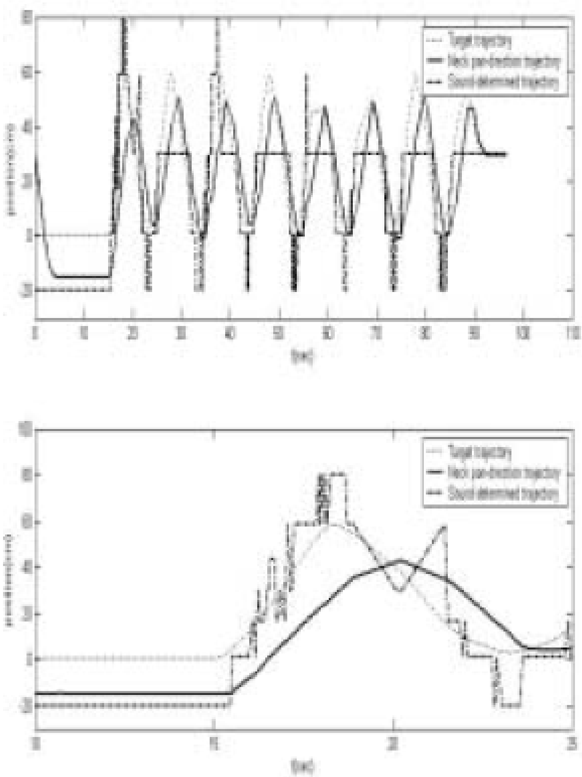

Fig. 6 Neck pan-direction trajectory

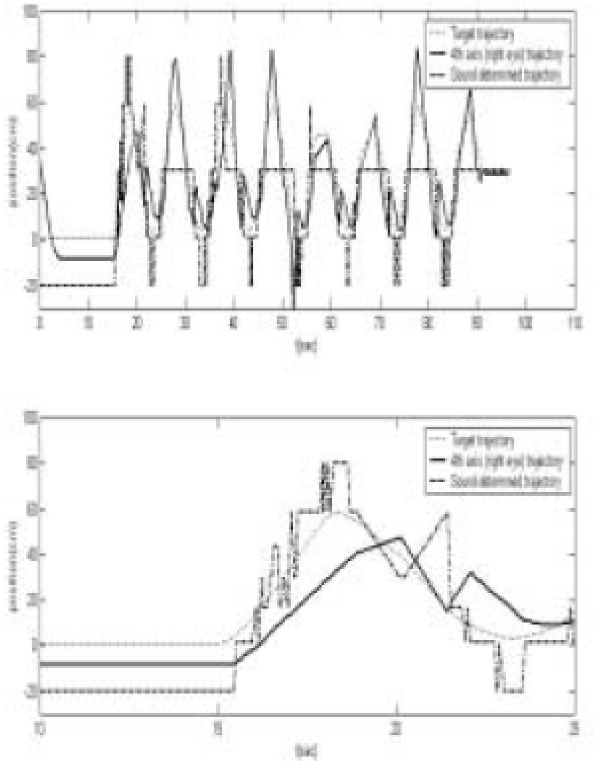

Fig. 7 Right eye optical axis trajectory

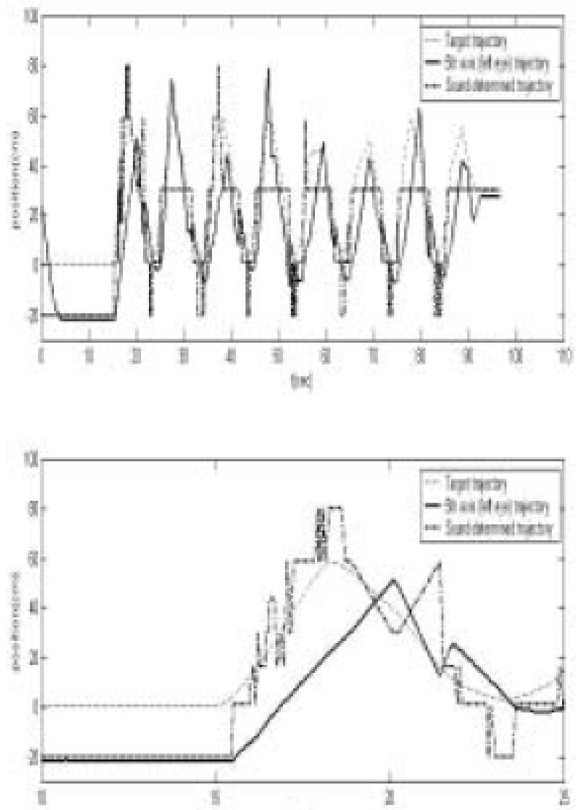

Fig. 8 Left eye optical axis trajectory

## V. CONCLUSION

A robust binocular tracking system with an auditory perception system mounted on the top of the robot is developed. The advantages of this tracking system are as follows.

1. An auditory perception system is formed from an equilateral triangular array of microphones.
2. A more efficient image tracking algorithm is proposed, based on blob determination. The edge-blobs pattern matching distinguishes the internal characteristics of the target from the other background objects, and edge-blobs contour matching distinguishes the target's external characteristics. These two new methods reduce the number of matching below that associated with full-range searching method or the 3SHS matching method.
3. YCbCr is applied to construct a color tracking space. It filters out background noise whose color is not similar to that of the target. Therefore, visual tracking works in complicated circumstances.
4. With the aid of the auditory perception system, the robotic head can be driven to search the covered target, just as a human can hear a voice while not being able to see the speaker.

## REFERENCES

[1] P. Y. Chen (2003), A Robust Visual Servo System for Tracking an Arbitrary-Shaped Object by a New Active Contour Method. Master Thesis, Department of Electronical Engineering, National Taiwan University.

[2] G. K. Wang (2004), Design and Implementation of a Multi-Purpose Real-time Visual Tracking System based on Modified Adaptive Background Subtraction and Multi-Cue Template Matching. Master Thesis, Department of Electrical Engineering, National Cheng Kung University.

[3] K. Yanagisawa, A. Ohya and S. Yuta (1995), An Operator Interface for an Autonomous Mobile Robot Using Whistle Sound and a Source Direction Detection System. Proceeding of 21th IEEE IECON, pp.1118-1123.

[4] J. L. Chen (2002), Development of a Sound Direction Detection System. Master Thesis, Department of Electrical and Control Engineering, National Chiao-Tung University.

[5] J. Hu, T. M. Su, C. C. Cheng, et al (2002), A self-calibrated Speaker Tracking System Using both Audio and Video Data. Proceedings of the 2002 IEEE International Conference on Control Applications.