Sensor and Data Systems, Audio-Assisted Cameras and Acoustic Doppler Sensors

Kaustubh Kalgaonkar
Georgia Insitute of Technology
Georgia, Atlanta
kaustubh@ece.gatech.edu

Paris Smaragdis, Bhiksha Raj
Mitsubishi Electric Research Labs
Cambridge MA
paris,bhiksha@merl.com

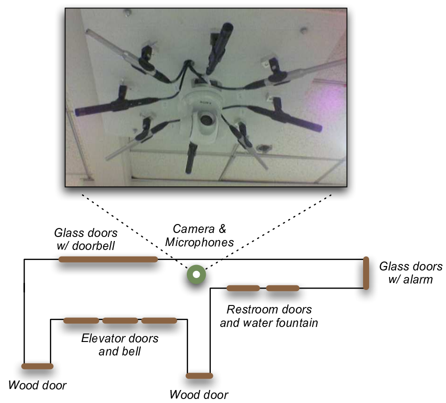

Figure 1. Prototype audio-assisted camera and its location

Abstract

In this chapter we present two technologies for sensing and surveillance – audio-assisted cameras and acoustic Doppler sensors for gait recognition.

1. Audio-assisted Cameras

Surveillance with cameras has become ubiquitous – most commercial locations are monitored by several cameras that constantly record all activity. However, the amount of data generated by such systems far exceeds human capacity to analyze it. In this section we describe our work on *audio-assisted* cameras, whereby, rather than indiscriminate recording of all activity, we direct PTZ cameras towards directions of interest, as determined by analysis of concomitant audio.

1.1. Prototype Setup

Figure 1 shows a prototype setup. A PTZ camera is augmented by an array of microphones. The camera is mounted on the ceiling of a high-activity area near an elevator. Our objective was to track people from all directions and detect and classify activity in the area.

1.2. Localization by Sound Recognition

We identify known sound sources, such as elevators, doors etc. (and their locations) using a Bayesian classifier. We represent incoming sounds through their magntigude STFTs, whose dimensions are reduced by PCA. We learn a mixture Gaussian density for the dimensionality-reduced spectra of each of the sound sources. During operation, we guide the camera in the direction of the most likely sound source.

1.3. Localization by Location Recognition

Where similar sound sources, e.g. doors, occur in multiple locations, they must be disambiguated. For each putative direction, for each pair of microphones, we learn a Gaussian density for the vector (component-wise) *ratio* of the magnitude spectra of the signals captured. We augment these with the distribution of the vector difference of the phase spectra of the two signals. Since the latter wraps around at 2π, the distribution must be modelled with a modified Gaussian density learned through an EM algorithm [1]. Incoming signals are now classified as arriving from the direction with the greatest likelihood. If this likelihood falls below a threshold, we back off to a more traditional method for localization [2].

1.4. Applications

The proposed audio-assisted surveillance has been successfully applied on traffic surveillance and accident detection on data from the Japan national police archives. We have also used it effectively for elevator surveillance, monitoring and tracking human traffic, etc.

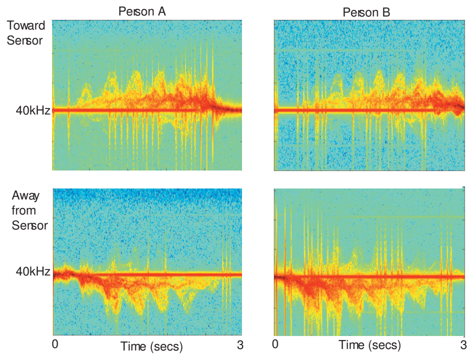

Figure 2. Spectrograms of Ultrasonic reflections for 2 walkers

2. Gait Recognition with Acoustic Doppler

"Gait" is defined as a person's manner of walking. A person's gait often carries information beyond that present in the mere physiognomy of the individual – it is often possible to identify a person from a distance by their gait even when they are too far away to distinguish their facial or other features. In her seminal studies, Murray [3] proposed that the totality of a person's gait, including the entire cycle of motion, is a unique and identifying characteristic.

Traditionally, the identification of gait has been considered a visual phenomenon – it is a characteristic of a person that must be seen to be recognized. We present an *acoustic* Doppler based gait recognition system that uses frequency shifts of low-frequency ultrasonic tones to characterize motion. We demonstrate that the proposed mechanism can provide highly accurate gait recognition, even using only simple signal processing schemes that are conventionally employed for analysing audio data, and an equally simple Bayesian classifier.

2.1. Acoustic Doppler Sonar for Gait Recognition

A human body is an articulated object, comprising a number of rigid bones connected by joints. When a continuous tone is incident on a walking person, the reflected signal contains a spectrum of frequencies arising from the Doppler shifts of the carrier tone by moving body parts. Figure 2 show the spectrogram of reflections of a 40kHz tone from two subjects. It may be observed in Figure 2 that the patterns in the left and right panels, which represent the two subjects, are different, although superficially both exhibit similar cyclic patterns. It is these differences that enables us to recognize the walker using the reflected signals.

For our experiments we use a device similar to the one used by Kalgaonkar and Raj[4]. It comprises a 40kHz ultrasonic transmitter, driven by a 40kHz oscillator, and a corresponding 40kHz sensor (receiver) to capture reflected signals. The entire hardware costs less than $10 to build.

Experiment	No. evaluated	No. Correct	Percent correct
Walker identification	300	275	91.66
Approach vs. away	300	289	96.33
Male vs. Female	300	242	80.66

Table 1. Classification results for acoustic Doppler sensor

2.2. Signal Processing and Classification

The reflected signal sensed at the Doppler receiver is a sum of the signals reflected by the various moving components. To extract features from this signal we first demodulate it. If we represent the received signal as $d(t)$ and the carrier frequency by f_c, the demodulation operation is performed as:

$$y(t) = \mathrm{LPF}(\sin(2\pi f_c t)\frac{d}{dt}d(t)) \tag{1}$$

40-dimensional features are derived through cepstral analysis of short-time Fourier spectra of the demodulated signal. These are augmented by their 1^{st} to obtain a final 80-dimensional vector.

We modelled the distribution of features for each of the classes (subjects) through a simple Gaussian Mixture model (GMM), whose parameters were learnt from about 15 seconds of training recordings. The GMMs were then used to identify the walker using a standard Bayesian classifier.

2.3. Experiments and Results

Three types of classification experiments were carried out: walker identification, direction of walk (toward or away from sensor) and gender identification from gait. The results are presented in the Table 1.

References

[1] P. Smaragdis and P. Boufonos Position and trajectory learning for microphone arrays. *IEEE Trans. Audio, Speech and Language Processing*, January 2007. 1

[2] S.T. Birchfield and D.K. Gillmor Fast Bayesian acoustic localization. *Proc. IEEE ICASSP*, May 2002. 1

[3] M. Murray. Gait as a total pattern of movement. *American Journal of Physical Medicine*, 46(1):290–332, 1967. 2

[4] Kaustubh Kalgaonkar and Bhiksha Raj An acoustic Doppler-based front end for hands free spoken user interfaces. *Spoken Language Technology Workshop*, 46(1):158–161, 2006. 2