Applied Acoustics 129 (2018) 316–321

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Double mode surveillance system based on remote audio/video signals acquisition

Tao Lv [a,b], He-yong Zhang [a,*], Chun-hui Yan [a,b]

[a] Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, State Key Laboratory of Laser Interaction with Matter, Changchun 130033, China
[b] University of the Chinese Academy of Sciences, Beijing 10039, China

ARTICLE INFO

Article history:
Received 3 July 2017
Received in revised form 15 August 2017
Accepted 16 August 2017
Available online 20 August 2017

Keywords:
LDV
Laser hearing
Voice enhancement
Multimodal sensing

ABSTRACT

At present, remote human signature detection plays an increasingly important role in the field of anti-terrorism and security defense all around the world. In order to acquire remote, non-cooperative human signature signals, a double mode (audio/video) surveillance system is developed. The system mainly consists of a Laser Doppler Vibrometer (LDV), a pan-tilt-zoom (PTZ) camera and a theodolite. The LDV is used to acquire remote audio by detecting the vibration of the object (caused by the acoustic pressure). The PTZ camera is used to capture the video of the human body, and track the body when he/she moves, then analyze the image to select a proper vibrating target for LDV measurements. The theodolite is applied to control the orientation of the LDV. For the reason that many noise sources disturb the LDV-measured signals, such as laser speckle noises, environmental noises and the noises caused by sensor moving, a kind of speech enhancement technology (OM-LSA algorithm) is used to improve the intelligibility of the noisy voice signals detected by the LDV system. Experiments results indicated that the SNR and MOS of the LDV speech signal (the range is 150 m) can be increased by 135% and 26% respectively by using the speech enhancement technology, and the remote speech and video signals (the range is 30 m) can be obtained by the double-mode surveillance system.

1. Introduction

Remote human signature detection systems are widely deployed today for security purpose. In general, most remote human signature detection systems mainly depend on visual information [1]. Although video technologies (including visible and IR) have a great advancement in human signature detection, there is still a serious limitation in non-cooperative and hostile environments. The audio information, as an important data source, has not been fully explored yet. A few systems [2,3] have been reported to integrate visual and acoustic sensors, but in those systems, the acoustic sensors need to be close to the targets in monitoring. Parabolic microphones can be used for remote hearing and surveillance, which can capture voice at a fairly large distance in the direction pointed by the microphone. However, it is sensitive to the noise caused by wind and sensor motion. Laser Doppler Vibrometer (LDV) can measure extremely tiny vibration of a target at a long range [4–9]. On the other hand, objects near to the audio sources can be vibrated by the acoustic pressure. Therefore, the voice signals of a human could be acquired by capturing the vibration of a target's surface caused by the speech of the person next to the target. Li, Wang and et al [10–14] have presented their results in detecting and processing voice signals of people from large distances using a LDV from Polytec (includes a controller OFV-5000 with a digital velocity decode card VD-6, a sensor head OFV-505). However, in their work, the light of the LDV is 632 nm, which is not fit for realistic application due to its visible laser beam (can be perceived easily). Besides, the system is bulky and heavy because of the separated structure (the Polytec OFV 505 system has a size of 120 mm × 80 mm × 345 mm and weight of 3.4 kg).

In previous work, we have presented our results in detecting and processing voice signals of people from large distances using a self-designed LDV [15]. Now, a LDV with a configuration of all fiber is established. The all-fiber LDV system can offer the advantages of smaller size, lightweight design, robust structure, and the laser beam is invisible (1550 nm). Therefore, it is a better choice for remote speech detection. In this paper, the double-mode surveillance system based on self-design LDV, PTZ camera and theodolite is developed. The system can be used to detect remote human signature (visual and audio information). Several experiments are implemented to test its performance.

* Corresponding author.
E-mail address: zhanghy@ciomp.ac.cn (H.-y. Zhang).

http://dx.doi.org/10.1016/j.apacoust.2017.08.016

2. Experimental setup

The principle block diagram of double-mode surveillance system is shown in Fig. 1. This system is composed of LDV, camera, theodolite and a personal computer.

The LDV is composed of transceiver (Fig. 2) and signal processor (Fig. 3). A 20-mW single-frequency (1550 nm) fiber laser with single longitudinal mode and narrow linewidth (less than 10 kHz) is used as the transmitter. The beam from the laser is divided into two beams by a 1×2 fiber coupler, one part is used for the local oscillator (LO) beam, and the other part acted as the transmitted beam. In order to discriminate the direction of target vibration, an acousto-optic frequency shifter (AOFS) is equipped to the LO beam. Afterwards, the LO is frequency-shifted up by 40 MHz using the AOFS, whose driving signal served as the reference signal input to the signal processing unit. The transmitted beam is focused on the target after passing through a circulator and a telescope. Due to the vibration of the target (vibration caused by the voice energy), the reflected beam carries Doppler frequency shift. This reflected beam is received by the same telescope, and it is mixed with LO by a 2×2 fiber coupler. Finally, the interference signal is detected by a balanced photodetector.

The output of the balanced photodetector is an FM signal with a center frequency f_{AOFS} of 40 MHz. In order to obtain acoustic signal, the demodulation methods are needed (Fig. 3). The detector output signal and reference signal are sampled by a dual-channel high-speed (250 M/s) data acquisition card. Then the detector output signal is divided into two equal parts, and the two divided signals mix with two orthogonal replicas of the reference signal, the corresponding in-phase (I) and the quadrature (Q) output signals can be obtained after through the low-pass filters. Finally, the arctangent phase function reconstructed the audio signals. Besides, the

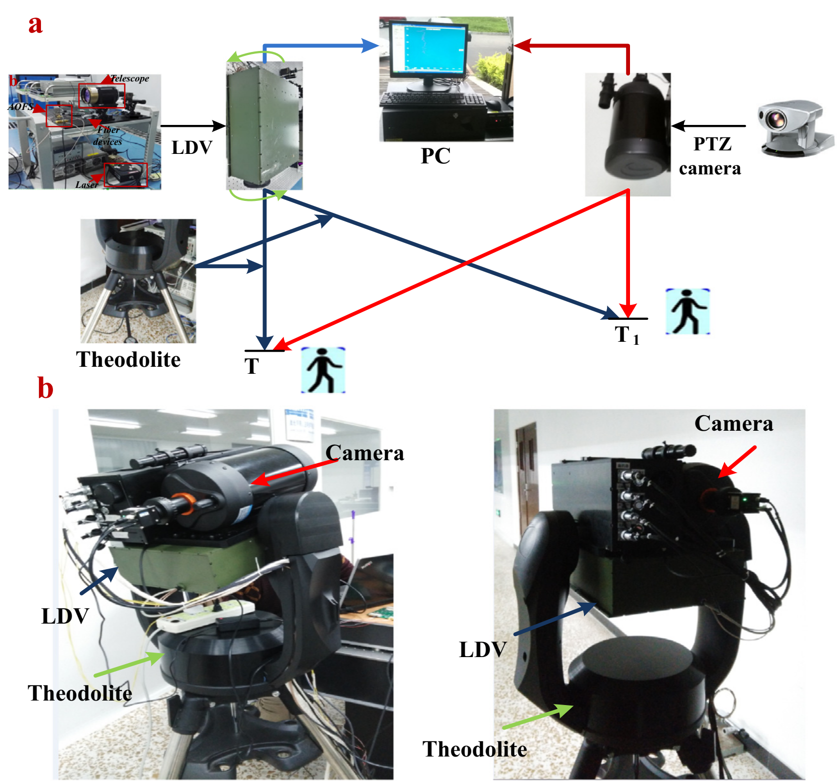

Fig. 1. (a) Schematic setup of the double-mode surveillance system. (b) The double-mode surveillance system.

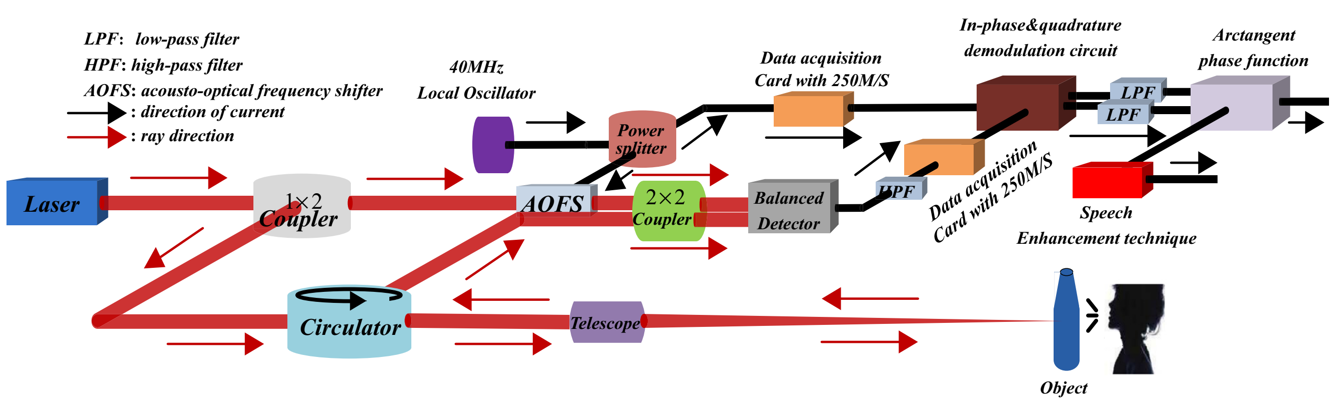

Fig. 2. Schematic diagram of the LDV transceiver.