# LDV Sensing and Processing for Remote Hearing in a Multimodal Surveillance System

Zhigang Zhu*, Weihong Li, Edgardo Molina and George Wolberg
Computer Science Department, CUNY City College and Graduate Center
Convent Avenue and 138th Street,New York, NY 10031
*zhu@cs.ccny.cuny.edu

## 1. Introduction

Recent improvements in laser vibrometry [1-3] and day/night infrared (IR) and electro-optical (EO) imaging technology [4, 5] have created the opportunity to create a long-range multimodal surveillance system. This multimodal capability would greatly improve security force performance through clandestine listening of targets that are probing or penetrating a perimeter defense. This system could also provide the feeds for advanced face and voice recognition systems.

The studies of the capabilities of these three types of sensors are critical to such surveillance tasks. IR and EO cameras have been studied and widely used in human and vehicle detection in traffic and surveillance applications [5]. Laser Doppler vibrometers (LDV) such as those manufactured by Polytec™ [1] and B&K Ometron [2] can effectively detect vibration within two hundred meters with a sensitivity on the order of 1μm/s. These instruments have been used to measure the vibrations of civil structures like high-rise buildings, bridges, towers, etc. at distances of up to 200m. However, literature on remote acoustic detection using the emerging LDVs is rare. Therefore, we mainly focus on the experimental study of the LDV-based voice detection, in the context of a multimodal surveillance system. We have also set up a system with the three types of sensors for performing integration of multimodal sensors in human signature detection.

We first give an overall picture of our technical approach: the integration of laser Doppler vibrometry and IR /color imaging for multimodal surveillance. One of the important issues is how to use IR and/or color imaging to help the laser Doppler vibrometer to select the appropriate targets. Then, we discuss various aspects of LDVs for voice detection. Acoustic signals captured by laser vibrometers need special treatment since the detected speech signals may be corrupted by more than one noise source, such as laser photon noises, target movements, and background acoustic noises (wind, engine sound, etc.). Finally, we provide a brief discussion on some future work in LDV sensor improvements and multimodal human signature detections.

## 2. Multimodal Sensors for Remote Hearing

There are two main components in our approach of multimodal human signature detection (Figure 1): the IR/EO imaging video component, and the LDV audio component. Both the IR/EO and LDV sensing components can support day and night operation even though it will be better to use a standard EO camera (coupled with the IR camera) to perform the surveillance task during daytime.

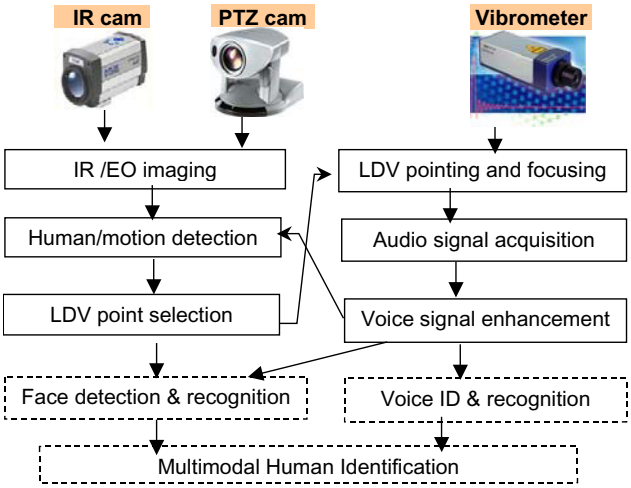


Figure 1. System Diagram.

The overall approach is the integration of the IR/EO imaging and LDV audio detection for a long-range surveillance task, which has the following three steps.

*Step 1. Target detection, tracking, and selection via the IR/EO imaging module.* The targets of interest could be humans or vehicles (driven by humans). This could be achieved by motion detection and human/ vehicle identification, either manually or automatically.

*Step 2. Audio targeting and detection by the LDV audio module.* The audio signals could be human voices or vehicle engine sounds. We mainly consider the human voice detection. The main issue is to select LDV targeting points provided by the IR/EO imaging module, not necessarily parts of the subjects, in order to detect the

1-4244-1180-7/07/$25.00 ©2007 IEEE

vibration caused by the subjects' voices.

*Step3. Face/vehicle shots of best views triggered by the feedback from audio detection.* By using the audio feedback, the IR/EO imaging module can verify the existence of humans and capture the best face shots for face recognition. Together with the voice recognition module, the surveillance system could further perform human identification and event understanding.

Figure 1 shows the system diagram. In our current experiments, the first step was performed manually. Note that a large body of automated algorithms exists in literature [e.g., 5] that can be readily applied here. The third step has also been very well-studied, including face detection/recognition, audio/video speech recognition and multimodal biometrics. The focus of the paper will be on the second step, including LDV target detection, LDV pointing, LDV voice detection and signal enhancements.

## 3. LDV Hearing: Sensing and Processing

Laser Doppler vibrometers (LDVs) work according to the principles of laser interferometry. Measurements are made at the point where the laser beam strikes the structure under vibration. There are two important issues to consider in order to use an LDV to detect the vibration of a target caused by human voices. First, the target vibrates with the voices. Second, points on the surface of the target where the laser beam hits reflect the laser beam back to the LDV. For perimeter surveillance, we use existing facilities or install special facilities for human audio signal detection. Facilities like walls, pillars, lamp posts, large bulletin boards, and traffic signs vibrate very well with human voices [6]. Note that the LDV has a sensitivity on the order of 1 μm/s, and can therefore pick up very small vibrations. We have found that most objects vibrate with voices, and many types of surfaces reflect the LDV laser beam within some distance (about 10 meters), without retro-reflective treatment. Response is significantly improved if we can paint or paste certain points of the facilities with retro-reflective tapes or paints; operating distances can increase to 300 meters (about 1000 feet) or more.

For the human voice, the frequency range is about 300 Hz to 3 KHz. However, the frequency response range of the LDV is much wider than that. Even if we have used the on-board digital filters, we still get signals that include troublesome large, slowly varying components corresponding to the slow but significant background vibrations of the targets. The magnitudes of the meaningful acoustic signals are relatively small, adding on top of the low frequency vibration signals. This prevents the intelligibility of the acoustic signals by human ears. On the other hand, the inherent "speckle pattern" problem on a normal "rough" surface and the occlusion of the LDV laser beam (by passing-by objects) introduce noises with large and high-frequency components into the LDV measurements. Therefore, we have applied Gaussian bandpass filtering [6] and Wiener filtering [7] to process the vibration signals captured by the LDV. In addition, the volumes of the voice signals may change dramatically with the changes of the vibration magnitudes of the target due to the changes of speech loudness (shouting, normal speaking, whispering) and the distances of the human speakers to the target. Therefore, we have also designed an adaptive volume function [6] to cope with this problem.

However, without retro-reflective tape treatment, the LDV voice signals are very noisy from targets at medium and large distances. Therefore, further LDV sensor improvement is required, including laser power, wavelengths and reflectance [6]. With current state-of-the-art sensor technology, we realize that more advanced signal enhancement techniques need to be developed than the simple band-pass filtering, Wiener filtering and adaptive volume scaling. For example, model-based voice signal enhancement could be a solution in that background noises might be captured and analyzed, and models could be developed from the resulting data.

We also want to emphasize that automatic targeting and intelligent refocusing is one of the important technical issues that deserve attention for long-range LDV listening, since it is extremely difficult to aim a laser beam at a distant target and keep it focused. We believe that LDV voice detection techniques combined with the IR/EO video processing techniques can provide a more useful and powerful *remote* surveillance technology for both military and civilian applications.

This work is supported by AFRL contract F33615-03-1-6383, NSF Grant No. CNS-0551598, and the CUNY Equipment Grant Competition Program.

## References

[1] Polytec Laser Vibrometer, http://www.polytec.com/
[2] Ometron, http://www.imageautomation.com/
[3] MetroLaser, http://www.metrolaserinc.com/vibrometer.htm
[4] FLIR Systems. `http://www.flir.com/`
[5] R. Hammoud, *Joint IEEE International Workshops on Object Tracking and Classification Beyond the Visible Spectrum (OTCBVS)*, 2004, 2005, and 2006.
[6] Z. Zhu, W. Li and G. Wolberg, Integrating LDV audio and IR video for remote multimodal surveillance, *OTCBVS'05.* See also an extended TR CUNY TR-2005006 in HTML at http://www-cs.ccny.cuny.edu/~zhu/LDV/FinalReportsHTML/CCNY-LDV-Tech-Report-html.htm
[7] W. Li, M. Liu, Z. Zhu and T. Huang, LDV Remote Voice Acquisition and Enhancement, *ICPR'06*. Suppl. material: http://www-cs.ccny.cuny.edu/~zhu/LDV/ICPR06-LDV-Suppl.rar