Active Stereo Vision for Improving Long Range Hearing Using a Laser Doppler Vibrometer

Tao Wang[1], Rui Li[1,2], Zhigang Zhu[1], Yufu Qu[1,2]

[1] Dept. of Computer Science, City College of City University of New York, USA
Email: {twang, zhu}@cs.ccny.cuny.edu
[2] School of Instrumentation Science & Opto-electronics, Beihang University, China

Abstract

Laser Doppler Vibrometers (LDVs) have been widely applied for detecting vibrations in applications such as mechanics, bridge inspection, biometrics, as well as long-range surveillance in which acoustic signatures can be obtained at a large distance. However, in both industrial and scientific applications, the LDVs are manually controlled in surface selection, laser focusing, and acoustic acquisition. In this paper, we propose an active stereo vision approach to facilitate fast and automated laser pointing and tracking for long-range LDV hearing. The system contains: 1) a mirror on a Pan-Tilt-Unit (PTU) to reflect the laser beam to any locations freely and quickly, and 2) two Pan-Tilt-Zoom (PTZ) cameras, one of which is mounted on the Pan-Tilt-Unit (PTU) and aligned with the laser beam synchronously. The distance measurement using the stereo vision system as well as triangulation between camera and the LDV laser beam allow us to fast focus the laser beam on selected surfaces and to obtain acoustic signals up to 200 meters in real time. We present some promising results with the collaborative visual and LDV measurements for laser pointing and focusing in order to achieve long range audio detection.

1. Introduction

Recently, audio-visual sensory systems have been used for biometrics [1], activity recognition [2], and large area surveillance [3, 4]. Typically, in these systems, the visual sensors are Pan-Tilt-Zoom (PTZ) cameras and the acoustic sensors are microphones. However, for long range object detection, microphones need to be placed close to the targets of interest in order to obtain their acoustic signatures. Many microphones need to be fixed at pre-determined locations if an object is in motion and needed to be tracked. Parabolic microphones, which can capture voice signals at a fairly large distance in the direction pointed by the microphone, could be used for remote hearing and surveillance. But it is very sensitive to noise caused by the wind or the sensor motion, and all the signals on the way are captured. A Laser Doppler Vibrometer (LDV) is a novel type of measurement device to detect a target's vibration in a non-contact way, in applications such as bridge inspection [5], biometrics [6], and underwater communication [7]. It has also been used to obtain the acoustic signals of a target (e.g., a human or a vehicle t) in a large distance by detecting the vibration of a reflecting surface caused by the sound of the target next to it [8, 9].

However, in most of the current applications, such systems are manually operated. In close-range and lab environments this is not a very serious problem. But in field applications, such as bridge/building inspection, area protection or search and rescue applications, the manual process takes a very long time to find an appropriate reflective surface, focus the laser beam and get a vibration signal; more so if the surface is at a distance of 100 meters or more. For example, using a COTS Polytec LDV (OFV-505), it takes about 15 seconds for the laser focusing on a surface. In our previous work [10], a vision-aided LDV system is proposed to improve the performance and the efficiency of the LDV for automatic remote hearing. The platform integrated a PTZ camera, a mirror, and a PTU to the LDV. The PTZ camera not only assists the LDV to point the laser beam to a better reflective surface for obtaining audio signals, but also measures the distance of the target by using triangulation of the PTZ camera, the mirror, and the LDV. However, the design and system has some limitations. First, since the laser cannot be pointed to the mirror at the center of the rotation of the PTU, the calibration procedure is sophisticated, and the laser pointing for finding a new surface via the PTZ camera is not very straightforward due to the fact that there is not a common center of projection for the laser beams controlled by the PTU. Second, there is the chicken or the egg problem in the laser focusing and laser point detection. In order to focus the laser quickly on a target surface using the distance measurement from triangulation, the laser spot has to be seen and detected in the PTZ camera images, which is hard when the laser beam is not focused. Finally, there is also a tradeoff between the 3D accuracy and visibility. In order to obtain sufficient and accurate distance measurement over a large distance (>100 meters), the baseline between the camera and the mirrored LDV system has to be large. However, this will increase the difficulty in coordinating the two sensors to point the laser beam to a designated surface location; in the worst case, the laser spot might be occluded by some closer objects and therefore is not in the field of view of the PTZ camera.

In this work, we improved flexibility and usability of our previous by using a pair of PTZ cameras, one of which is mounted on the top of the Pan-Tilt-Unit (PTU). This camera will be called the master camera and the other the slave camera. Furthermore, a mirror is mounted on the rotation center of the same PTU so that the laser beam from the LDV is always pointing to the rotation center; by controlling the PTU, the LDV can point to any directions quickly and freely, meanwhile, the viewing direction of the PTZ on top of the PTU (the master camera) always align with the direction for the reflected laser beam. Therefore even though the laser spot cannot be seen (and detected) before it is focused, its location is known in the master camera. By finding correspondence of the location into the slave camera, the distance of that location can be determined and the laser can be focused. In this work, we will present how we obtain the target distance initially using the stereo vision technique to fast focus the laser spot. Then, the laser beam is moved automatically through control of the mirror and the master camera. The returned laser signals and acoustic signatures allow us to verify the correctness of distance measurements of the two PTZ cameras.

PTZ camera networks and stereo vision using PTZ cameras have been widely used and well studied for wide-area video surveillance [11-14]. The main contribution of this paper is the collaborative operation of a dual-PTZ-camera system and a laser pointing system for long-range acoustic detection. To our knowledge, this is the first work that uses a PTZ stereo for automating the long-range laser-based voice detection. Meanwhile, the combination is a natural extension of the already widely used PTZ-camera-based video surveillance system towards multimodal surveillance with audio, video and range information.

The rest of the paper is organized as follows. Section 2 presents an overview of our audio-visual sensory system for long range detection. Section 3 describes the stereo match techniques for the two cameras and with the LDV. Section 4 discusses the adaptive and cooperative sensing between two modalities. Section 5 provides some experimental results. Finally, we conclude our work in Section 6.

2. System Overview

The system consists of a single point LDV, a mirror mounted on a PTU, and a pair of PTZ cameras with one mounted on the top of the PTU (Fig. 1).

The Polytec LDV sensor OFV-505 and the controller OFV-5000 we use in our experiments can be configured to detect vibrations under several different velocity ranges: 1 mm/s/V, 2 mm/s/V, 10 mm/s/V, and 50 mm/s/V, where V stands for velocity. For voice vibration of a basic frequency range from 300 to 3000 Hz, we usually use the 1mm/s/V velocity range. The best resolution is 0.02 μm/s under the range of 1mm/s/V according to the manufacture's specification with retro-reflective tape treatment. Without the treatment, the LDV still has sensitivity on the order of 1.0 μm/s. This indicates that the LDV can detect vibration (due to voice waves) at a magnitude in nanometers without retro-reflective treatment or even picometer with retro-reflective treatment.

Figure 1. The multimodal sensory platform

The LDV sensor head weights about 3.4 kg; this is the major reason that a mirror mounted on the PTU is employed in our system to reflect the laser beam to freely and quickly point to directions of a large field of view. The laser beam point to the mirror at the center of the panning tilting of the PTU. The vision component consists of two Canon VC-C50i (26x) PTZ cameras with one mounted on the top of the PTU, which is called master PTZ since it is the main camera to track the laser beam, and another one mounted on the top of the LDV, which is called slave PTZ. The reason to use zoom cameras is to detect targets and to assist the laser pointing and focusing at various distance. However, at a long distance, the laser spot is usually hard to be seen by the either zoomed or wide views from either of the PTZ cameras if the laser is unfocused or not pointed on the right surface. Therefore, the master PTZ camera is used to rotate synchronously with the reflected laser beam from the mirror in order to track the laser spot. Although the laser point may not be observed from the master PTZ, we always control the pan and tilt of the master PTZ camera so that its optical axis is in parallel to the reflected laser beam and therefore the laser spot is always close to the center of the image. Then, the master PTZ camera and the slave PTZ form a stereo vision system to obtain the distance to focus the laser spot as well as guide the laser to the right surface for acoustic signal collection. The baseline between of the two PTZ cameras is about 0.6 meters for enabling long range distance measurements. In order to obtain the distance from the target surface to the LDV, the calibration among the two PTZ cameras and the LDV is first important step, which we will elaborate in the following before we discuss our method in distance measurement.

3. System Calibration and Stereo Matching

There are two stereo components in our system: stereo vision between the two PTZ cameras, and stereo triangulation between the slave PTZ camera and the mirrored LDV laser projection. The first component is used to obtain the range of a point in a reflective surface by matching its image projection (x, y) in the master camera to the corresponding image point (x', y') in the salve camera. The second component is mainly used to obtain the pan (α) and tilt (β) rotation angles of the PTU so that the LDV points to the image point (x, y) in the master image. Before determining the distance, several coordinate systems corresponding to the multi-sensory platform (in Fig. 1) is illustrated in Fig. 2 in a left-side view: the master PTZ camera coordinate system (S_c), the slave PTZ camera coordinate system ($S_{c'}$), the LDV coordinate system (S_L), and the PTU coordinate system (S_u).

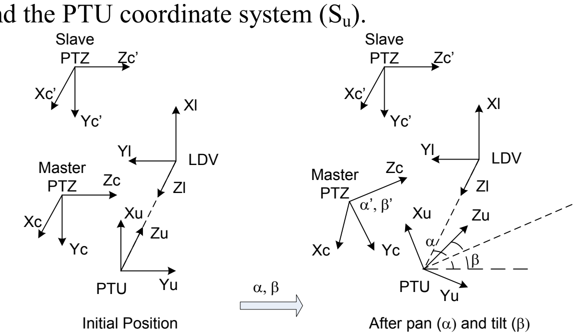

Figure 2. Platform coordinate systems

We assume the mirror coordinate system is same as the PTU coordinate system since the laser will point to the mirror at the origin of the PTU system. The mirror normal direction is along the Z_u axis and initially points to the outgoing laser beam along Z_l. In order to always track the reflected laser beam by having the optical axis of the master PTZ parallel to the reflected laser beam, the master PTZ not only rotates the same angles with the PTU, α and β, which are the pan and tilt angles of the PTU around the X_u and Z_u axes, but also undergo additional pan and tilt rotations (α' and β') around the Y_c and X_c axes. We will explain in details how to determine these angles in Section 4.

The stereo matching is performed after the full calibration of the stereo component of the two PTZ cameras, and that between the slave PTZ camera and the "mirrored" LDV. Given a selected point on a reflective surface in the image of the master camera, we first find its corresponding point in the image of the slave camera, meanwhile calculating the pan and tilt angles of the PTU and the master and slave PTZ camera so that the laser spot is right under the center of the image of the master PTZ camera; the offset to the center is a function of the distance of the surface to the sensor system. The farther the surface is, the closer is the laser spot to the center. The distance from the target point to the optical center of the LDV is estimated via the stereo PTZ and then used to focus the laser beam to the surface.

3.1. Calibration of the two PTZ cameras

The calibration between the two PTZ cameras is carried out by estimating both the intrinsic and extrinsic parameters of each camera on every possible zoom when the camera is in focus, using the same world reference system. We use the calibration toolbox by Bouguet [15] to find a camera's parameters under different zoom factors. We have found that the estimated extrinsic parameters do not change much with the changes of zooms. However, the focal lengths of the cameras increase nonlinearly with the changes of different zooms (Fig. 3) therefore we have calibrated the camera under every possible zoom. Also note that the focal lengths of two PTZ cameras may not be same under the same zoom factor. In order to achieve similar FOVs and to ease the stereo matching between two images, the correct zoom of the slave PTZ camera corresponding to the actual focal length of the master PTZ camera should be selected. After the calibration, we obtain the effective focal lengths and image centers of the two cameras under every zoom factor k, and the transformation between the two cameras, represented by R and T:

$$P_{C'} = RP_c + T \qquad (1)$$

where P_c and $P_{c'}$ are the representations of a 3D point in the master and slave PTZ coordinate systems (S_c and $S_{c'}$), respectively.

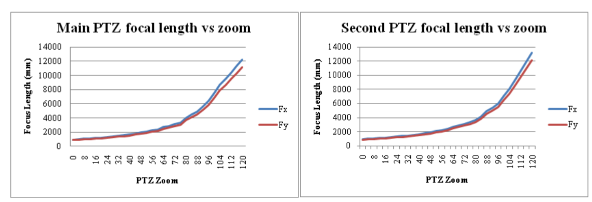

Figure 3. Focal lengths vs. zoom factors of both PTZ cameras

3.2. Calibration of the slave camera and the LDV

Since the intrinsic parameters of the slave PTZ camera have been obtained previously, we only need to estimate the extrinsic parameters characterizing the relation between the LDV coordinate system (S_L) and the slave PTZ camera coordinate system ($S_{C'}$), defined as:

$$P_L = R_{C'} P_{C'} + T_{C'} \qquad (2)$$

where P_L and $P_{C'}$ represent the coordinates of a 3D point in S_L and $S_{C'}$, respectively. The $R_{C'}$ and $T_{C'}$ are the rotation matrix and translation vector between S_L and $S_{C'}$. Next, the relation between the LDV coordinate system (S_L) and the mirrored LDV coordinate system (S_{ML}, not shown in Fig. 1) is defined as:

$$P_L = R_U R_{LR} R_U^T (P_{ML} - T_U) + T_U \qquad (3)$$

where P_L and P_{ML} are the 3D point representations in the S_L and the S_{ML}, respectively, and R_U and T_U are the rotation matrix and translation vector between S_L and the PTU coordinate system. The R_{LR} is the rotation matrix that converts a right hand coordinate system to a left hand coordinate system. Then the extrinsic parameters are estimated by combining Eq. (2) and Eq. (3), as

$$R_{C'}P_{C'} = R_U R_{LR} R_U^T (P_{ML} - T_U) + (T_U - T_{C'}) \quad (4)$$

For the calibration between the LDV and the slave PTZ, the LDV laser beam is projected at pre-selected points in a checkerboard placed at various locations/orientations. Because both the variables P_{ML} - T_U and T_U -$T_{C'}$ are not independent in Eq. (4), the distance between the fore lens of the LDV and the laser point on the mirror is estimated initially. Also, to avoid the complexity of the nonlinear equation we assume the initial rotation matrix is identify matrix which can be manually adjusted by pointing both cameras parallel to the same direction. Then this initial distance and initial rotation matrix can be refined iteratively. Giving n 3D points, 3n linear equations that include n+14 unknowns are constructed using Eq. (4). Therefore, at least 7 points are needed.

3.3. Distance measurements

After calibration, distance of a point can be estimated when the corresponding point in the slave image of a selected point in the master image is obtained. In the master camera, a target point can be selected either manually or automatically. For example, in Fig. 4, the right image is captured by the master PTZ camera and the left image by the slave PTZ camera. The same target points are shown in white circle. The epipolar line (using Eq. (1) is shown in green line cross both images. Note that due to the calibration error, the corresponding point may not be exactly on the epipolar line. To solve the problem, a small search window is used to match the region around the selected point with a range of vertical shift as well. Since we are only interested in the selected point on a particular reflective surface, this method is feasible and fast.

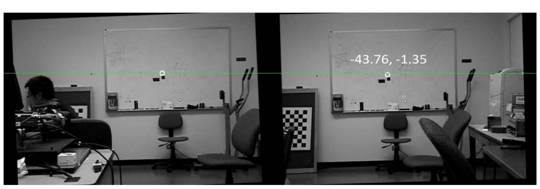

Figure 4. Stereo matching of the same target point. The white numbers on the right image show the pan and tilt angles of the PTU in order to point the laser beam to the target point.

The distance (D) and the 3D coordinates $P_{c'}$ =(X', Y', Z'=D) of the target surface represented in the slave camera system can be easily obtained using triangulation between the two cameras (Eq. (1)). Then the calibration result of the slave camera and the LDV is used to determine the pan (α) and tilt (β) angels of the PTU to point the laser beam to the selected point. The conventional triangulation method is used to match the ray from the optical center of the PTZ to the ray of the reflected laser beam, and the corresponding pan and tilt angels are estimated. Fig. 4 shows an example of the calculated pan and tilt angels (on the right image) corresponding to the point (in white circle) in the left image. Then, the estimated distance from the LDV center to the point can be calculated using on Eq. (4). We have the LDV distance D_L = $||P_{ML}||$ due to that the Y and Z coordinates of the laser beam are zeros in the mirrored LDV coordinate system. Note that in all these calculations, we have put the new rotations of the two PTZ cameras and the PTU in their relations, in Eq. (1) and Eq (4). This distance value will be used for focusing the laser beam to the target point.

4. Adaptive and Collaborative Sensing

The overall goal of this system is to acquire useful audio signatures with the assistance of video cameras by pointing and focusing the laser beam to a good surface. However, a target location either manually or automatically selected may not return a good signal. A reselection of new target points is required. Fig. 5 shows the basic idea of adaptive sensing of adaptively adjust the laser beam based on the feedback of its returned signal levels.

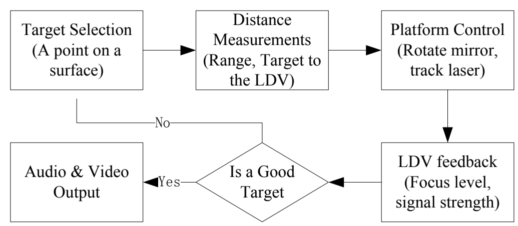

Figure 5. Flow chart of adaptive sensing for laser pointing and tracking for audio and video signature acquisition

The stereo matching here is used for obtained the target distance to the system platform, then we can automatically focus the laser point to the selected target. This involves the following procedures. First, a point on a surface close to a designated target is selected either manually or automatically. Second, the target range and the distance from the point to the optical center of the LDV are measured. Third, the laser spot is moved to the new location, and the master PTZ camera is rotated synchronously to put the laser spot in the center of images. Fourth, the laser beam of the LDV is automatically and rapidly focused based on estimated distance and the signal levels, as we did in [10]. As a summary, once the estimated distance is obtained, we search a range of focus steps of the LDV and select the focus step yield the strongest returning signal. If the selected target point does not have sufficient good returning signals for voice detection, we need to reselect new target points. If the target point is good enough, we can use it to record the audio signature as well as video signatures (e.g., faces) if we are dealing with human subjects.. Usually, we need to select several points around the human object and choose the best one.

There is a detailed technical issue in tracking the laser spot in step 3, especially for long range detection. The laser spot may not be observable at a long range particularly if it is not focused or it does not point on the surface accurately. We solve this problem by keeping the reflected laser beam always in parallel to the optical axis of the maser PTZ camera so that the laser spot is right under and very close to the center of the master image (Fig. 6). We make the ray from the optical center of the master PTZ camera parallel to the reflected laser beam by rotating the PTZ camera with the PTU synchronously (since the PTZ is mounted on the PTU), then with additional PTZ camera rotations (explained below).

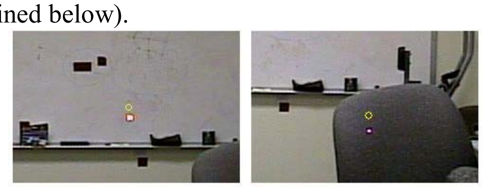

Figure 6. Two examples of laser point tracking. Both images show the same cropped size (240x160) around the image center (yellow circle) with focused laser spot close to it (red-white dot).

The main issue now is how to obtain the additional pan (α') and tilt (β') angles of the PTZ camera given the pan (α) and tilt (β) angles of the PTU. Fig. 7 shows the relationship between the outgoing laser beam from the LDV ($\overrightarrow{BA}$) and the reflected laser ray ($\overrightarrow{AD}$), with the mirror normal ($\overrightarrow{AC}$). By projecting the reflected ray and the mirror normal on the YZ plane (in both the PTU and the LDV coordinate systems in Fig. 2), as $\overrightarrow{AE}$ and $\overrightarrow{AF}$ respectively, we see that the angle ∠*BAF* is α and the angle ∠*FAC* is β. Now the camera optical axis is parallel to the mirror normal AC. Two additional angles are defined in the figure, the pan angle α' as the angle ∠*FAE*, and the tilt angle β' as the angle ∠*EAD*. If the master PTZ camera (mounted on top of the PTU) is tilted back by - β, then its optical axis will be parallel to AF. Therefore, by further panning the PTZ by the angle α' and tilting the PTZ by the angle β', its optical axis will be in parallel with the reflected laser ray AD.

Define a helping line GC parallel to AD, we can easily solve the additional pan (α') and tilt (β') based on triangulation. The detailed derivation is straightforward and is neglected here. As a result, the pan angle (α') is

$$\alpha' = \tan^{-1}\left(\frac{\tan\alpha}{\cos 2\beta}\right) \tag{5}$$

and the tilt angle (β')

$$\beta' = \sin^{-1}(\sin 2\beta * \cos\alpha) \tag{6}$$

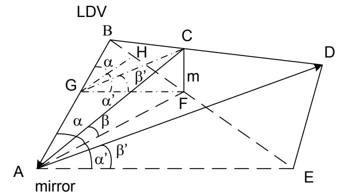

Figure 7. Geometry model of laser beam from the LDV ($\overrightarrow{\boldsymbol{BA}}$) and its reflected laser ray ($\overrightarrow{\boldsymbol{AD}}$) after the pan (α) and tilt (β).

5. Experimental Results

We test the system under different environments, inside our lab (with distances up to 10 meters), in the corridor (up to 31 meters), and on the street (up to 100 meters). Each experiment involves four parts: 1) verify the accuracy of the calibration of the two PTZ cameras, 2) verify the calibration accuracy of the slave camera and the LDV, 3) test the correctness of laser point tracking using the collaboration of the master camera and the PTU+mirror platform, and 4) demonstrate the idea of adaptive sensing based on the laser returned signals. Here, we only show the experiment in the corridor. Several interested target points are automatically selected in the segmented regions close to the human target in an image of the maser PTZ (Fig. 8).

Figure 8. Interested target points close to the human target in the segmented regions are selected and labeled. Original image of the corridor (about 31 meters) under zoom factor 48 with a human target in the center (in red ellipse), is shown as an inset.

To verify the correctness of calibration parameters, especially with change of the focal lengths of each camera, we use the same feature point on the check board pattern at various distances. At each zoom level, the distance from the check board to the platform is manually obtained as the ground truth, and then we use the calibrated parameters to estimate the distance at that zoom level. Fig. 9 shows the comparison of the true and estimated distances under various zoom factors, which has an average relative error of 6%. The accuracy is sufficient for performing the adaptive focus of the LDV sensor.

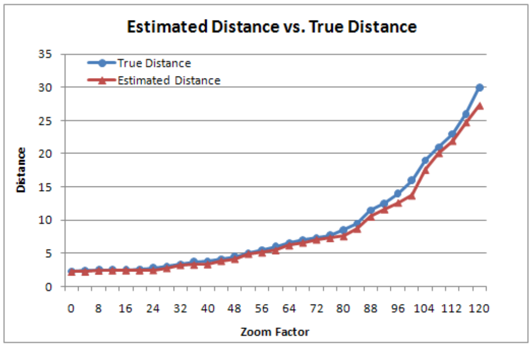

Figure 9. The comparison of true distance and estimated distance under various zoom factors.

Table 1. Surface selection, laser pointing and focusing

L#	Surface		Measurements			Ground Truth		
		D (m)	D_L(m)	Step	Level	D* (m)	Step	Level
L1	Wall	29.67	30.26	2786	12	30.63	2845	14
L2	Mirror	27.90	28.56	2758	12	30.63	2757	12
L3	Metal box	29.67	30.26	2745	18	30.32	2839	21
L4	Side wall	21.10	21.60	2391	9	23.16	2410	11
L5	Wall	30.85	31.43	2410	11	30.63	2757	11
L6	Wall	28.62	29.20	2734	11	30.63	2734	12
L7	Wall	27.67	28.25	2732	12	30.63	2734	12
L8	Chalkboard	28.30	28.88	2750	11	27.74	2764	12
L9	Woodside	26.56	27.14	2642	12	27.74	2581	12

The experimental results related to parts 2, 3 and 4 for those labeled points are presented in Table 1. The estimated camera distance (D) are listed in column 3 with the "ground truth" data (D*) at column 6. The LDV distance (D_L) in column 4, the distance from the target point to the optical center of the LDV, is calculated based on the pan and tilt angles of the PTU. Base on that, the focus step (in the range of 0 to 3300) in column 5 is determined and the laser beam is focused in about 1 second for each point. For comparison, the focus step using the full range searching takes 15 second, and is presented in column 8. The signal returning levels (0 to 512) in column 6 can be used to determine what the best point is among the candidates for audio acquisition. As a result, the metal box (L3) has the strongest signal return level so that it is selected for the voice detection.

6. Conclusion

In this paper, we present a dual-PTZ camera based stereo vision system for improving the automation and time efficiency of LDV long-range remote hearing. The close-loop adaptive sensing using the multimodal platform allows us to determine good surface points and to quickly focus the laser beam based on target detection, surface point selection, distance measurements, and LDV signal return feedback. The integrated system greatly increases the performance of the LDV remote hearing and therefore its feasibility for audio-visual surveillance and long-range other inspection and detection applications.

Acknowledgments

This work is supported in part by AFOSR under Award #FA9550-08-1-0199, NCIIA under an E-Team Award #6629-09, ARO -DURIP under Award # W911NF-08-1-0531 and by NSF under grant No. CNS-0551598. We thank Prof. Jizhong Xiao for his assistance in the prototyping of the hardware platform.

References

[1] W. Chen, S. B. Oetomo, L.Feijis, P. Andriessen, F. Kimman, M. Gerates, M. Thielen, Rhythm of life aid (ROLA): an integrated sensor system for supporting medical staff during cardiopulmonary resuscitation (CPR) of newborn infant," IEEE Trans. Information Technology in Biomedicine, issue 99, May 2010.

[2] T. Petsatodis, A. Pnevmatikavakis, C. Boukis, Voice activity detection using audio-visual information, 16th Int. Conf. Digital Signal Processing, August, 2009.

[3] M. Cristani, M. Bicego and V. Murino, Audio-visual event recognition in surveillance video sequences, IEEE Trans. Multimedia, 9(2): 257-267, February. 2007

[4] Y. Dedeoglu, B. U. Toreyin, U. Gudukbay and A. E. Cetin, Surveillance using both video and audio, in Multimodal Processing and Interaction: Audio, Video, Text, P. Maragos, A. Potamianos and P. Gros Eds., 2008: 143-156.

[5] A. Z. Khan, A.B. Stanbridge and D.J. Ewins, Detecting damage in vibrating structures with a scanning LDV, Optics and Lasers in Engineering, 32(6), 1999:583-592.

[6] P. Lai, et al, A robust feature selection method for noncontact biometrics based on laser Doppler vibrometry, " IEEE Biometrics Symposium, 2008: 65-70

[7] F. A. Blackmon and L. T. Antonelli, Experimental detection and reception performance for uplink underwater acoustic communication using a remote, in-air, acousto-optic sensor, IEEE J. Oceanic Engineering, 31(1), Jan. 2006: 179-187

[8] W. Li, M. Liu, Z. Zhu and T. S. Huang, LDV remote voice acquisition and enhancement, ICPR 2006: 262-265.

[9] Z. Zhu, W. Li and G. Wolberg, Integrating LDV audio and IR video for remote multimodal surveillance, OTCBVS'05.

[10] Y. Qu, T. Wang and Z. Zhu, An Active Multimodal Sensing Platform for Remote Voice Detection, IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM 2010), Montreal, July 6-9, 2010, pp. 627-632..

[11] J. Zhou, D. Wan and Y. Wu, The chameleon-like vision system, IEEE Signal Processing Magazine, 27(5), Sept, 2010: 91-101.

[12] S.-N. Lim, A. Elgammal, L.S. Davis, Image-based pan-tilt camera control in a multi-camera surveillance environment, ICME '03, vol 1: 645-648

[13] E.D. Nelson and J.C. Cockburn, Dual camera zoom control: a study of zoom tracking stability, ICASSP 2007: I-941 - I-944

[14] A. W. Senior, A. Hampapur, M. Lu, Acquiring multi-scale images by pan-tilt-zoom control and automatic multi-camera calibration, WACV'05: 433-438

[15] J. Y. Bouguet, Camera calibration toolbox for Matlab. http://www.vision.caltech.edu/bouguetj/calib_doc/index.html, June 2008