MODAL-BASED CAMERA CORRECTION FOR LARGE BASELINE STEREO IMAGING

Nathan Short*, Prather Lanier, Kevin Kochersberger**, and Lynn Abbott***
*Bradley Dept. of Electrical and Computer Engineering
** Unmanned Systems Laboratory, Dept. of Mechanical Engineering
Virginia Tech, Blacksburg, VA 24061
{nshort21, abbott, pjlanier, kbk}@vt.edu

ABSTRACT

This paper is concerned with stereo imaging for three-dimensional range estimation. The usual assumption for a two-camera system is that both cameras are stationary with respect to one another. In some imaging environments, however, physical vibrations are unavoidable. If these vibrations induce small movements of the cameras relative to each other, then the accuracy of any range estimates based on binocular disparity will suffer. These problems are especially pronounced for mobile, large-baseline systems such as aerial vehicles. This paper describes an approach for reducing loss of accuracy by modeling vibrational camera motion using external sensors such as accelerometers, and then triggering image acquisition to coincide with known orientations of the cameras. Experiments using two different camera platforms are presented. The first is a simple beam that is center-mounted to a shaker to induce symmetric bending. For the second set of experiments, cameras were mounted near the wingtips of an unmanned aerial vehicle (UAV). In both cases, ranging accuracy was significantly improved when vibrational movements were incorporated into the system.

INTRODUCTION

Stereopsis relies on images that are obtained from slightly different points of view. Because a single point in the three-dimensional (3-D) scene projects onto different relative locations in the images, the disparity between corresponding points can be used to estimate the 3-D location of the original point. A common approach is to use two cameras that are triggered simultaneously. A search for correspondences in the two images produces a 3-D point cloud representing surface points for physical objects in the scene.

The accuracy of the estimated point locations depends on a calibration procedure that provides knowledge of the positions and orientations of the cameras relative to each other, along with intrinsic parameters such as focal lengths of the lenses. Typically, it is assumed that both cameras are rigidly mounted, so that the relative positions and orientations do not change. In such a case, the calibrated parameters can be assumed to be constant. This assumption is reasonable for systems with small pitch (baseline) distances separate the two cameras. For example, commercially available systems are rarely found with a pitch exceeding 0.6 m.

For some applications, however, larger baselines may be desirable. One reason is that larger camera separations allow for higher resolution in the range estimates. Unfortunately, as larger baselines are employed, it becomes more difficult to prevent relative movements of the cameras. This is particularly true for mobile platforms, such as aircraft, where light-weight designs are needed. For fixed-wing aircraft, it is often desirable to mount cameras near the wingtips in order to obtain large baselines without a significant amount of additional weight. Such solutions unfortunately introduce relative camera movements as the wings flex.

This paper describes a novel approach to addressing problems of accuracy that result from nonrigid stereo camera mounts. We make the assumption that the relative motion between two cameras is due to vibrations that can be measured and modeled. If the vibration modes are known, then it becomes possible to make small corrections in the calibrated camera orientations based on the times that images were acquired. Our earlier work (Lanier, Short, Abbott, & Kochersberger, 2009) demonstrated that these small corrections can improve accuracy when cameras are mounted on a flexible beam, such as the system shown in Figure 1. Using vibrational measurements from accelerometers, modal analysis was used to predict camera movements and to improve the accuracy of range estimates. This paper extends the previous work by demonstrating the validity of this approach for more general camera mounts, such as airframes.

STEREO VISION

A stereo vision system utilizes two or more cameras that are positioned to capture images of the same 3-D scene from different points of view. This camera placement causes slight differences in appearance in the images, and these differences (known as stereo disparities) can be used to extract 3-D information from the 2-D projections. A common imaging arrangement is motivated by biological vision, with cameras that are horizontally aligned and separated by a small baseline distance. The stereo system used in the initial experiments is shown in Figure 1.

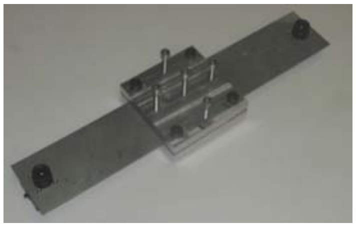

Figure 1. Stereo cameras mounted on a flexible beam with a 10 inch baseline.
The two lenses are visible near the ends of the metal bar.

Stereo ranging is illustrated with the simple arrangement that is shown in Figure 2. In this ideal system, the optical axes of the two cameras are perfectly parallel, both image planes are coplanar, and no lens distortion is present. Scene point *P* projects onto both image planes, and we would like to recover its 3-D coordinates. In this case, the distance *Z* (also called the range or depth) can be found using the following:

$$Z = \frac{Bf}{d} \qquad (1)$$

In this equation, *f* is the focal length, *B* is the baseline, and *d* is the disparity between two corresponding points in the left and right images, which is given by $d = x^l - x^r$. (In the figure, the value for x^r is negative because it is on the left of the optical axis.) Although stereo ranging is simple in principle, the identification of point correspondences between the two images is a difficult problem.

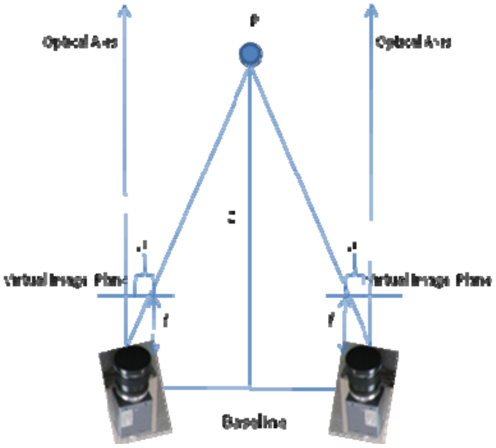

Figure 2. Simple geometry for stereo ranging. The usual goal is to find the range *Z* from the cameras to a point *P* in the scene.

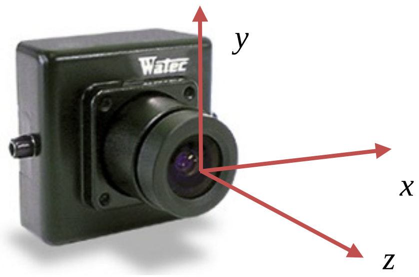

Figure 3. Camera-centered coordinate system. This shows a Watec 660D G3.8 mono board camera, which is the model used for this project.

In an actual stereo system, the optical axes are not perfectly parallel. A calibration procedure such as the one described in (Short, 2009) is used to determine the relative 3-D orientation between the two cameras, as well as intrinsic parameters such as the focal lengths and image center locations. With this information it is possible to rectify the two images with 2-D transformations that cause all corresponding points to be aligned horizontally. Rectification causes the images to resemble the ideal case shown in Figure 2, and it increases the efficiency of the search for corresponding points.

A rotation matrix can describe the orientation of the left camera with respect to the right, using the camera-centered coordinate system shown in Figure 3. The origin is located at the point of projection, and the *z* axis coincides with the optical axis. The *x* and *y* axes are parallel to the image plane. Rotation about the *x, y*, and *z* axes will be represented as θ_x (pitch), θ_y (yaw), and θ_z (roll) of the camera, respectively.

The composite rotation vector *RV* for one camera can be expressed as the row vector

$$RV = [\theta_x, \theta_y, \theta_z] \tag{2}$$

where $\Theta_x, \Theta_y, \Theta_z$ are the rotation vectors of the image plane. If *RV* represents a fixed axis through the origin, then the angle of rotation about this axis is given by the vector norm and the corresponding normalized unit rotation vector is Ω , with components $\Omega = [\Omega_x, \Omega_y, \Omega_z]$. If we now define the antisymmetric matrix, Ω_v, then the rotation matrix is given by R .

$$\theta - \|RV\| \tag{3}$$

$$\Omega = \frac{RV}{\theta} \tag{4}$$

$$\Omega_v = \begin{bmatrix} 0 & -\Omega_z & \Omega_y \\ \Omega_z & 0 & -\Omega_x \\ -\Omega_y & \Omega_x & 0 \end{bmatrix} \tag{5}$$

$$R - I + \Omega_v \sin\theta + \Omega_v \Omega_v^T (1 - \cos[\theta]) \tag{6}$$

where *I* is the 3×3 identity matrix. This is known as Rodrigues' rotation formula, and it can be rewritten as

$$R = [\blacksquare(\cos\theta + [\Omega_{vx}]^2 (1 - \cos[\theta]) \,\&\, -\Omega_{vz} \sin\theta + \Omega_{vx}\Omega_{vy} (1 - \cos[\theta]) \,\&\Omega_{vy} \sin\theta + \Omega_{vx}\Omega_{v} \tag{7}$$

For both cameras, the combined rotation matrix is

$$R = R_r R_l^T \tag{8}$$

where R_r and R_l are the rotation matrices for the right and left cameras, respectively.

In a typical stereo ranging system, it is assumed that the cameras will remain stationary relative to one another. After completing the calibration procedure, any change in relative pose between the two cameras will affect the stereo disparity values of imaged points, and will therefore reduce the accuracy of range estimates. If the system is subjected to vibrations that can affect the relative camera orientations, however, then it is expected that accuracy will suffer. Vibration can also introduce motion blur into the images, which affects the ability to localize feature points in the images and can further reduce ranging accuracy.

This paper considers situations in which two cameras are separated by a relatively large baseline distance, using a structural support that can flex when the system is subjected to vibration. The primary concern is rotation by the cameras in opposite directions about their respective *y* axes. Vibration will tend to cause the cameras to converge (rotate toward each other) and diverge (rotate away from each other) repeatedly. At maximum deflection during convergence, the disparity values for corresponding points will be smaller than for the calibrated (stationary) system. Based on the analysis given above, the estimated range values would therefore be smaller than the true values. Conversely, at maximum deflection during divergence, the disparity values will be larger, and the estimated range values would be larger than the calibrated values. Ironically, maximum deflection corresponds to minimum camera motion, which is the best time to capture images in an effort to reduce the effects of motion blur. Accelerometers are attached to the cameras for estimating deflection angles and to correct for small rotational changes of the cameras. Assuming periodic motion, the accelerometers can also be used to predict the instants of maximum deflection, which is when image actuation should be triggered to reduce blur.

MODAL MODELING OF A CENTRALLY-SUPPORTED BEAM

Several assumptions are needed for the development of the vibration model. Following from the development given in (Lanier, Short, Abbott, & Kochersberger, 2009), the assumption is that the cameras are centrally supported. For example, our cameras could be placed along the wings of an airplane with the fuselage acting as a mounting point. Next, we assume that this central support creates a fixed end condition on either side with the beam on which the cameras are attached. The beam, or hypothetical wing, behaves as an Euler-Bernoulli beam with tip mass. Additionally, the cameras will be mounted on the elastic axis of the beam to eliminate torsional moments from transverse deflections. Under these constraints, the system will behave as cantilever beam in transverse vibration.

The governing equation of vibration for the Euler-Bernoulli beam is, (Inman, 2001):

$$\frac{\partial^2 Y(x,t)}{\partial t^2} + c^2\left(\frac{\partial^4 Y(x,t)}{\partial x^4}\right) = 0 \qquad c = \sqrt{\frac{EI}{\rho A}} \tag{9}$$

where A is the cross sectional area of the cantilever beam, E is the elastic modulus, ρ is the beam density, and I is the inertial cross-section. The solution of the spatial equation has the form of $Ae^{\sigma x}$, and the general solution of the spatial equation, where $Y_n(x)$ is the cantilever beam mode shape, is

$$Y_n(x) = c_1 \sin(\beta_n x) + c_2 \cos(\beta_n x) + c_3 \sinh(\beta_n x) + c_4 \cosh(\beta_n x) \tag{10}$$

In this equation, c_n values are the modal constants that will be determined by the boundary conditions of a cantilever beam with a tip mass, which are as follows:

$$\textit{Fixed End Bondary conditions } (x = 0) \qquad Y = 0 \quad \frac{dY}{dx} = 0 \tag{11}$$

$$\textit{Free End Bondary conditions } (x = l) \qquad EI\left(\frac{\partial^2 Y}{\partial x^2}\right) = 0 \quad \frac{\partial}{\partial x}\left(EI\frac{\partial^2 Y}{\partial x^2}\right) = -m\omega^{\square}(x,t) \tag{12}$$

Here, Y is the analytical beam deflection and m is the mass of the stereovision camera. These boundary conditions assume that the tip mass (camera) has a very small rotary inertia compared to the beam. Differentiating equation 17 and applying the appropriate boundary conditions results in:

[∎(0&1&0 @1&0&1@ − sin(βl)& − cos(βl)&sinh(βl)@ − 【EIβ^3 cos (βl) + mω^2 sin(βl)& 【EIβ^3 sin】 (βl) (13)

For a non-zero solution, the leading matrix determinant is set to zero which provides the roots of the system. These roots can be used to determine the natural frequencies of our test camera boom. Using this approach, the fundamental bending frequencies of the system are found. The results are listed below.

After obtaining the natural frequencies of our stereo boom, the modes of vibration are determined by substituting the βl values, the weighted beam frequency values, where *l* is the length of the beam. Figure 4 shows the first mode shape (displacement) along with the derivative of this shape to obtain the modal slopes. It will be the slope plot that is used to correct the camera pose so that correct distances are obtained from the stereo image pair.

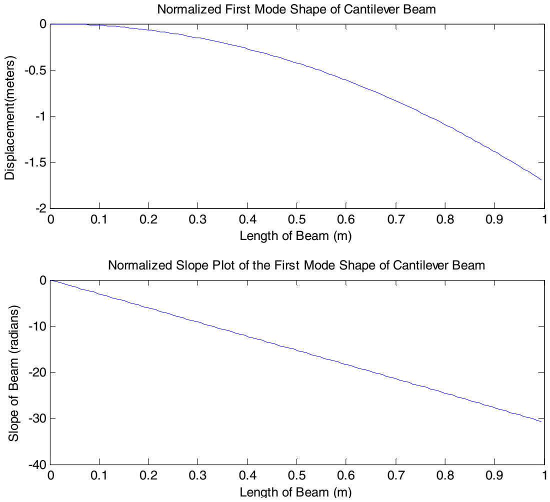

Figure 4. Normalized First Mode Shape of Stereo boom (Top). Normalized Slope along the length of stereo boom (Bottom).

Table 1. Physical properties of beam

Beam Properties	
Young's Modulus,(E)	$70 * 10^9\ GPa$
Density,(ρ)	$2700 \frac{kg}{m^3}$
Inertia,(I)	$4.225 * 10^{-15}\ m^4$
Cross sectional Area,(A)	$3.07 * 10^{-8}\ m^2$
Mass of Tip Mass(m)	$.035\ g$
Length of Beam (l)	$.1115\ m$

Table 2. Natural frequencies of first three modes of cantilever beam system

Mode	βl Values	Frequency (Rad/s)	Frequency (Hz)
1	1.047	126.305	20.1028
2	3.974	1865.8	296.955
3	7.098	5952.3	947.34

EXTENSION TO UNMANNED AERIAL PLATFORM

Based on the success of the method for flexible cantilever beams, the modal-analysis approach was tested experimentally using a UAV platform. The wingspan of the platform was 2 meters, giving a baseline of 1.98 meters for the stereo system. This setup is shown in Figures 5, and 6. A modal analysis of the wing requires that it be suspended as to allow the structure to respond as it would in actual flight conditions. This setup will be suspended using a technique outlined in (Woehrle, Costerus, & Lee, 1994), called the bungee cord method, shown in Figure 5. A wooden frame was constructed to support a flying wing suspended by bungee cords at three points on the platform. This type of mount allows the observance of natural flight behavior of the wing. Accelerometer placement and analysis of accelerometer input is outlined in Sections 5 and 6 of (Lanier P. J., 2010).

Two cameras were mounted on the wing tips of a fixed wing UAV as shown in Figure 6. The cameras were calibrated to obtain a calibrated stereo vision system for distance measurement. The cameras used in this experiment were two monochrome Unibrain fire wire 640 x 480 resolution cameras with a maximum frame rate of 30 frames per second.

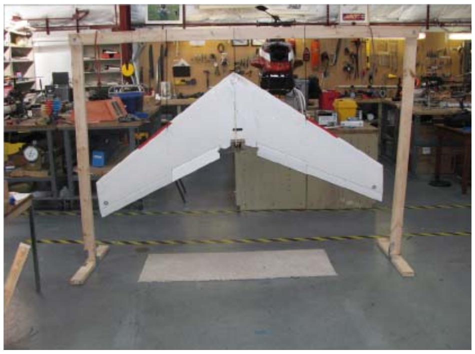

Figure 5. Experimental flying wing suspended from wooden frame with bungee cords.

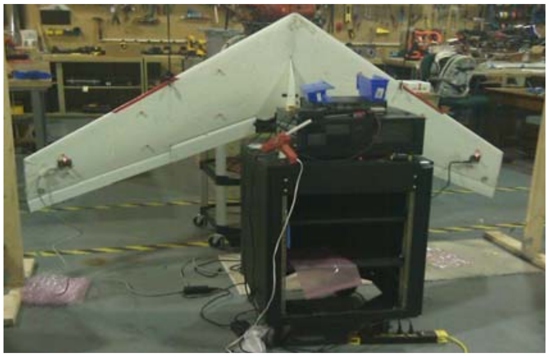

Figure 6. Front view of experimental setup with cameras mounted at wing tips.

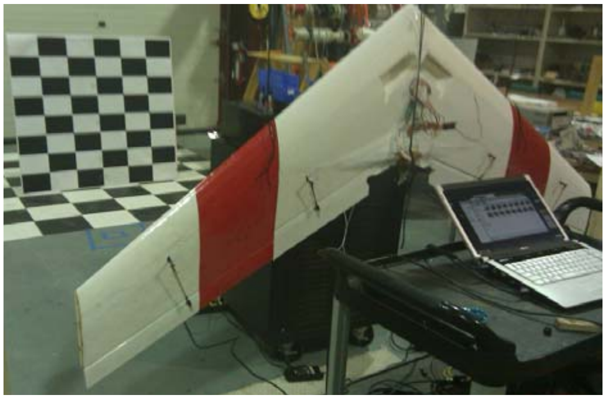

Figure 7. Reverse view of experimental setup with calibration target in view. Additional objects were placed in front of the system for distance measurements during the experiment.

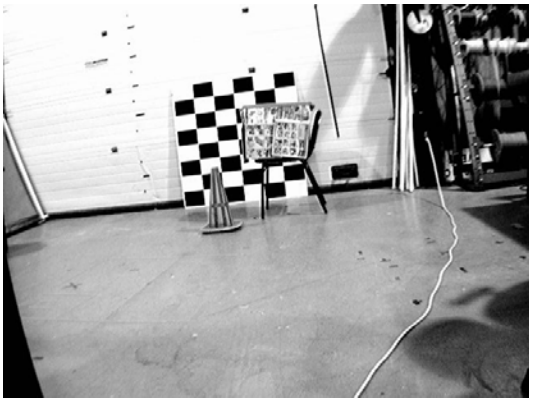

Figure 8. View of scene from right camera.

The Stereo system was calibrated using a calibration rig, Figure 7, and the same process that is outlined in (Short, 2009). After the system was calibrated, a test scene was set up for measuring distances using the images from the left and right camers, shown in Figure 8.

From the mode shape of the wing, determined from modal analysis as in (Lanier, Short, Abbott, & Kochersberger, 2009), the camera slope in relation to its position on the wing can be found. This is accomplished from two operations, scaling the approximate mode shape to the proper deflections at the camera position and taking the derivative of the scaled mode shape to calculate the values corresponding to the camera position. The polynomial fit of the mode shape of the wing is shown in Figure 9 with the wing deflection plot and slope plot in Figure 10.

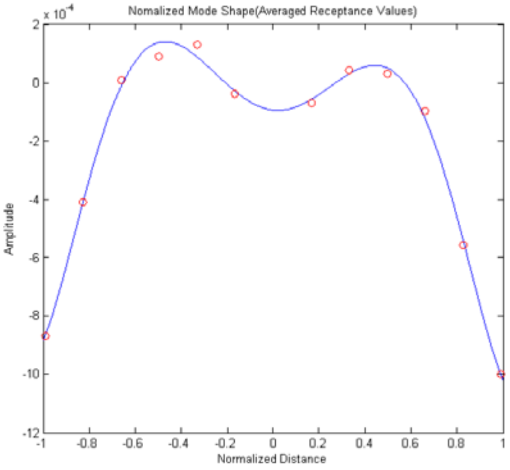

Figure 9. Polynomial fit of modal peaks across flying wing. Leading edge polynomial fit (top). Trailing edge polynomial fit (bottom).

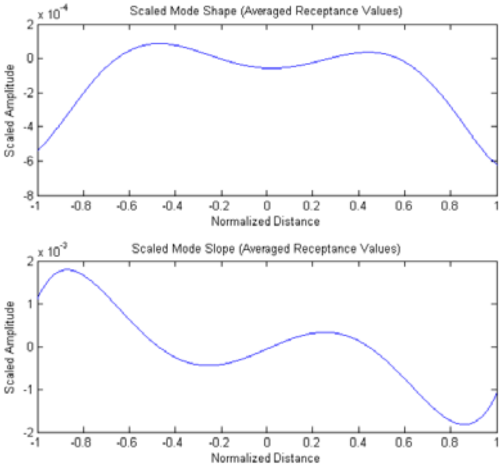

Figure 10. Mode shape that has been scaled to proper camera deflections (top). Slope of first experimental wing's first bending mode used to determine camera angle (bottom).

Table 3. Distance measurements from three objects to stereo system before vibration and after vibration without correction angle (top) and before vibration and after vibration with correction angle (bottom).

Object	Stereo System measurement without vibration	With Vibration without correction angle	% Error
Checker Board	5.137 m	5.078 m	1.14%
Bucket	4.096 m	4.051 m	1.09%
Cone	3.936 m	3.891 m	1.14%
		With Vibration with correction angle	
Checker Board	5.137 m	5.148	0.20%
Bucket	4.096 m	4.088	0.19%
Cone	3.936 m	3.936	~0.0%

Table 4. Scaled distance measurements from theoretical object at 60 meters to stereo system before vibration and after vibration without correction angle (top) and before vibration and after vibration with correction angle (bottom).

Object	Stereo System measurement without vibration	With Vibration without correction angle	% Error
Object	60 m	52.8 m	12.0 %
		With Vibration with correction angle	
Object	60 m	56.8 m	~5.3%

From the slope plot, we can find the angle of deflection of the cameras which provides the total correction angle to apply to the images in order to obtain improved distance measurements from the stereo system. The results before and after the correction angle is applied are shown in Table 3. As this experiment was performed in doors, the distance to the target was restricted, which explains the reason for such little error in measurements. A wide baseline system such as the one designed for this experiment, c. 2 meter baseline, is more suitable for long range stereo imaging for multiple reasons discussed in (Short, 2009). For this reason, scaled results were found based on the analysis of the same unmanned platform as they would appear from a distance of 60 meters, which is more of the range seen from a UAV using this wide platform. The error incurred from vibration is shown along with the corrected measurement found from applying the corrected angle. Using the formula in (1) and the fact that the difference in disparity at 60 meters from the deflection angle will be the same as it is at 5.13 meters, the error in distance measurement can be found. We can assume the same angle would be required, as the mode shape of the wing would remain the same.

As Table 4 shows, the error in distance measurement is greater due to the deflection at a greater distance. This is due to the decrease in resolution of the stereo system as the distance between the system and the object it is viewing increases. In this scenario, the distance measurement was corrected by 55.5% using the angle measured from the accelerometers and the plots from the modal analysis.

MODAL CORRECTIONS FOR SYSTEMS SUBJECT TO RANDOM NOISE

The flexible beam stereo vision system used in this experiment represents a system with inherent flexibility which must be characterized to obtain accurate stereo images. For this experiment, the beam and camera system represent a system with very small distributed mass along the beam and a large concentrated mass and inertia at the end of the beam. One of the reasons for this set-up was to create a low-frequency first mode for a small-pitch camera system; it would normally be considered bad practice to build a stereo vision system with such a flimsy backbone structure. For this experiment however, the system provided the low frequencies necessary for the data acquisition system to sync the camera images with the measured accelerations.

In practice, a properly designed system would have significant mass in the beam to keep the natural frequency high and the deflection low. One such example is in aircraft, where cameras mounted on the wingtips represent a wide baseline imaging system subject to dynamics that will require corrections for accurate stereo imaging. In this case, a modal analysis of the vehicle would be experimentally performed to determine how translational acceleration correlates to camera rotation, so camera corrections can be performed using accelerometer measurements.

For slender beam systems such as aircraft wings, gust conditions are expected to excite predominately the first mode wing bending. (Lee & Lee, 1993), and (Eslimy-Isfahany & Banerjee, 1997) have shown that second wing bending mode frequencies for aircraft are typically five times higher than the first mode frequency, and (Balakrishnan, 2006) has shown that turbulence energy for subsonic flight is concentrated at very low frequencies. (Bennet & Yntema, 1959) derived a method of calculating wing excitation from turbulence that only considers plunging rigid body motion and the first wing bending mode, reinforcing the concept that only the first flexible bending mode is typically excited in turbulence.

A simple demonstration was performed to show how the first wing bending mode is expected to dominate the flexible response of an aircraft in turbulence. Using the current set-up, random noise from 0 – 200 Hz was input to the shaker, and the response at the beam tip was measured for 100 averages. Figure 11 shows the random noise input (top) and the beam response (bottom), and Figure 12 shows the results in the frequency domain with a clear indication that the first mode is dominant. In practice, camera trigger would occur at the peak response, and first mode bending behavior would be used to correct the camera pose.

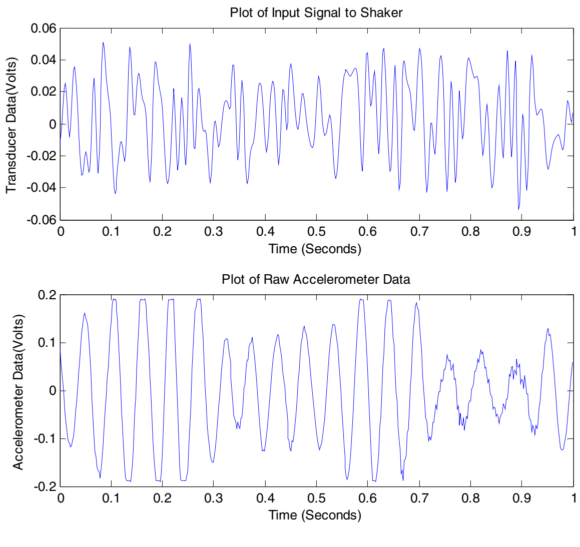

Figure 11. Input to Shaker (Top).Raw Data from Accelerometer (Bottom).

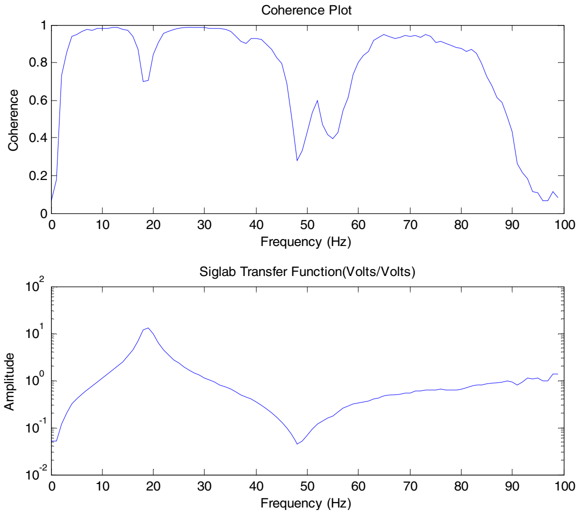

Figure 12. Coherence Plot (Top). Transfer Function (Bottom).

CONCLUSIONS

This paper has described a novel approach to stereo image acquisition in the presence of vibration. We found that modeling the stereo vision boom as a cantilever beam allowed for a reasonable prediction of the deflection angle of a stereo vision camera. It was shown in the first experiment of (Lanier, Short, Abbott, & Kochersberger, 2009), that the correction angle when used in the rectification of stereo images collected during steady-state vibration, the reduced the error in distance measurements by more than 70% when compared to their non-corrected counterparts. The results of the same approach when applied to an unmanned platform were also shown with an improvement in distance measurement of more than 50%. This represents a significant improvement in stereo ranging accuracy, and has potential for broad application in the field of unmanned systems.

REFERENCES

Axis Communications, 2009. *Axis Communications*, Retrieved September 28, 2009, from http://www.axis.com/products/cam_241q/index.htm

Balakrishnan, A. V., 2006. Modeline Response of Flexible High-Aspect-Ratio Wings to Wind Turbulence, *Journal of Aerospace Engineering* , Vol. 19, No. 2, p. 121 - 132.

Bennet, F. V., & Yntema, R. T., 1959. *The Evaluation of Several Approximate Methods for Calculating Symmetrical Bending-Moment Response of Flexible Airplanes to Isotropic Atmospheric Turbulence,* NASA TN 2018059L.

Bradski, G., & Adrian, K., 2008. *Learning OpenCV: Computer Vision with the OpenCV Library,* Sebastopol: O'Reilly Media Inc.

Eslimy-Isfahany, S. H., & Banerjee, J. R., 1997. Dynamic Response of Composite Beams with Application to Aircraft Wings, *Journal of Aircraft*, Vol. 34, No. 6, 785 - 791.

Inman, D. J., 2001. *Engineering Vibration,*Upper Saddle River, New Jersey: Prentice-Hall Inc.

Lanier, P. J., 2010. *Stereovision Correction Using Modal Analysis,* M.S. Thesis, Virginia Tech.

Lanier, P., Short, N., Abbott, L., & Kochersberger, K., 2009. Modal-based Camera Correction for Large Pitch Stereo Imaging, *Proceedings of the 24th International Modal Analysis Conference*.

Lee, I., & Lee, J. J., 1993. Vibration Analysis of Composite Wing and Tip Mass Using Finite Elements, *Computers and Structures*, Vol. 47, No. 3, p. 495-504.

Shapiro, L. G., & Stockman, G. C., 2001. *Computer Vision,* Upper Saddle River: Prentice-Hall, Inc.

Short, N. J., 2009. *3-D Point Cloud Generation from Rigid and Flexible Stereo Vision Systems,* M.S. Thesis, Virginia Tech.

Watec, 2008. *Watec Cameras,* Retrieved September 28, 2009, from http://www.wateccameras.com/

Woehrle, T., Costerus, B., & Lee, C., 1994. Modal Analysis of PATHFINDER Unmanned Air Vehicle, *Proceedings-SPIE The International Society for Optical Engineering*, 1687 - 1687.