Measuring seismic signals with airborne stereo cameras

Thomas Rapstine & Paul Sava, Center for Wave Phenomena, Colorado School of Mines

SUMMARY

Environmental impact and cost of acquiring land seismic data are major factors to consider when designing a seismic survey. We explore a way to quickly record land seismic data without disrupting the ground surface to reduce environmental impact and cost. We propose passively recording seismic data using a drone-borne stereo video camera system. Two cameras are used to monitor a textured ground surface as it is displaced according to a real-life earthquake signal. The recovered displacement measurements are represented as probability distribution functions (PDF's), which provide the uncertainty of our measurements. We conclude that reasonable, and currently available, cameras and drone systems can be used to measure seismic signals with associated uncertainties.

INTRODUCTION

Environmental impact, access, and acquisition cost are practical factors to consider when designing a geophysical survey. In environmentally sensitive areas, for example, it may not be possible to perform a ground-based survey without disrupting or damaging a fragile environment or ecosystem. These environmental concerns are accompanied by financial limitations. In practice, cost can be reduced by optimizing acquisition design parameters such as receiver spacing, seismic source geometry, source signal length, and frequency content. Technological advancements can also reduce costs of acquisition and potentially open new opportunities for exploration; an excellent example is distributed acoustic systems (DAS) (Bostick III, 2000; Daley et al., 2013; Mestayer et al., 2011; Mateeva et al., 2012, 2013, 2014). We advocate in this paper that drones and stereo vision also have the potential to impact seismic data acquisition, in principle allowing for airborne seismic surveying systems at high speed, low cost, and low environmental impact.

We advocate measuring distance variations to the ground as a function of time using stereo vision. Stereo vision has been used for a long time to passively measure distance using two images recorded from offset cameras (Quam et al., 1972). Our left and right eyes allow for depth perception, just as two images taken from offset cameras allow for a distance estimate. To perform stereo vision, we simultaneously record two digital images from calibrated cameras with a known offset. We then measure the lateral shift, or disparity, between the points as observed in the two offset cameras. As we show later, this disparity is inversely proportional to the distance from the camera to the ground points observed in the images. The stereo vision process is described in more detail in the Theory section.

In this paper, we demonstrate a method for sampling a seismic wavefield that has potential to reduce costs while providing exciting new opportunities for geophysical exploration. The method anticipates using drones and cameras as a means to acquire seismic data without contacting the ground surface. First, we discuss the underlying stereo vision theory used to determine ground displacement variations with time. Ground displacement values, with associated uncertainties, are recovered for a collection of points representing a vibrating ground surface. Lastly, using a realistic computer graphics simulation, we demonstrate the feasibility of applying stereo vision to recover realistic earthquake motion from videos acquired with airborne cameras.

THEORY

We aim to measure ground displacement, without touching the ground, while obtaining an uncertainty estimate of our measurement using stereo vision. The stereo vision theory we summarize is not new and is common in standard computer vision textbooks such as (Szeliski, 2011). Stereo videos are represented as a sequence of frames taken from two cameras (left and right). The left and right cameras are positioned in a specific geometry, shown in Figure 1, which allows one to deduce the distance to points viewed by the stereo system.

We begin by defining three coordinate frames representing the origin of the world, left camera, and right camera. The world coordinate frame is a reference in which we describe the position and orientation of the stereo cameras and of the ground. It is our goal to consistently represent a PDF of ground position in the world coordinate frame using images from a stereo system. In the world coordinate frame, we denote the world and left camera origin as ${}^{w}\mathbf{o}_{w} = [0,0,0]^{T}$ and ${}^{w}\mathbf{o}_{l} = [0,0,-h]^{T}$, respectively. The right camera origin is a placed a distance b away from the left camera origin noted by ${}^{l}\mathbf{o}_{r} = [b,0,0]^{T}$. We adopt a notation where preceding superscripts denote the coordinate frame of a point and subscripts denote the label of a point; a preceding w, l, and r denote the world, left camera, and right camera coordinate frames or labels, respectively. We represent transformations between an arbitrary coordinate system a to another coordinate system b using a 4×4 matrix ${}^{b}_{a}\mathbf{H}$. The matrix is obtained by augmenting the rotation matrix from coordinate system a to b, ${}^{b}_{a}\mathbf{R}$, with the origin of a in b, denoted ${}^{b}\mathbf{t}_{a,org}$

$$ {}^{b}_{a}\mathbf{H} = \begin{bmatrix} {}^{b}_{a}\mathbf{R} & {}^{b}\mathbf{t}_{a,org} \\ \mathbf{0} & 1 \end{bmatrix}. \tag{1} $$

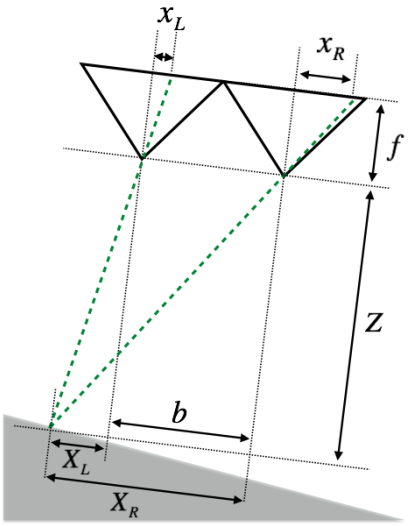


Figure 1: A schematic of stereo cameras viewing the ground from above. The two cameras have the same focal length f and optical axes that are aligned, but separated by a distance b. The dashed paths show where a point on the ground projects on the two cameras.

We use homogeneous point representation in order to apply point rotation and translation using a single matrix-vector multiplication. The homogeneous coordinates of a point may be obtained by appending the number 1 as a fourth element to the point's three-element vector representation. We represent a point $\mathbf{x}$ in homogeneous coordinates using the notation $\tilde{\mathbf{x}}$. We may obtain the original coordinate from the homogeneous coordinate by dividing by the fourth element

$$\tilde{\mathbf{x}} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \rightarrow \mathbf{x} = \begin{bmatrix} \frac{x_1}{x_4} \\ \frac{x_2}{x_4} \\ \frac{x_3}{x_4} \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}. \tag{2}$$

In general, a point in the right camera frame may be transformed into the world coordinate frame using

$${}^{w}\tilde{\mathbf{x}} = {}^{w}_{l}\mathbf{H}\,{}^{l}_{r}\mathbf{H}\,{}^{r}\tilde{\mathbf{x}} = {}^{w}_{r}\mathbf{H}\,{}^{r}\tilde{\mathbf{x}}. \tag{3}$$

We represent a point $\mathbf{p}$ in the left or right coordinate frames as ${}^{l}\mathbf{p} = [X_l, Y_l, Z_l]^T$ or ${}^{r}\mathbf{p} = [X_r, Y_r, Z_r]^T$, respectively. We note that $Z_l = Z_r = Z$ since the camera origins are only separated in the x-direction of the left camera frame. Using stereo vision, we find ${}^{r}\tilde{\mathbf{p}}$ then transform it to the world coordinates using ${}^{w}\tilde{\mathbf{p}} = {}^{w}_{r}\mathbf{H}\,{}^{r}\tilde{\mathbf{p}}$. The 3D location of the point ${}^{r}\mathbf{p}$ is found using the pixel coordinates where the point is observed in the left and right cameras, denoted in pixel coordinates as (x_l, y_l) and (x_r, y_r), respectively. Projecting a 3D point, such as ${}^{r}\mathbf{p}$, onto a 2D image can be described using a pinhole camera model. This model assumes a calibrated camera without lens distortions, implying that all 3D points project onto the image plane of a camera through a single point, or pinhole. Figure 1 shows a point projecting onto the left and right image planes using the pinhole camera model. We use similar triangle geometry to express the x-pixel coordinates of the point as a function of the camera focal length f, in pixels, and the distance to the point Z

$$\begin{aligned} \frac{x_l}{f} &= \frac{X_l}{Z} \rightarrow x_l = \frac{fX_l}{Z} \\ \frac{x_r}{f} &= \frac{X_r}{Z} \rightarrow x_r = \frac{fX_r}{Z}. \end{aligned} \tag{4}$$

The shift of the 3D point, as observed by the left and right cameras, known as disparity, is

$$d = x_r - x_l = \frac{fX_r - f(X_r - b)}{Z} = \frac{fb}{Z}. \tag{5}$$

We note that the shift is an integer since the pixel coordinates are integers. The distance, in the z-direction extending from the cameras, can be used to deduce the x and y-coordinates of ${}^{r}\tilde{\mathbf{p}}$

$$\begin{aligned} X_r &= \frac{Zx_r}{f} \\ Y_r &= \frac{Zy_r}{f}. \end{aligned} \tag{6}$$

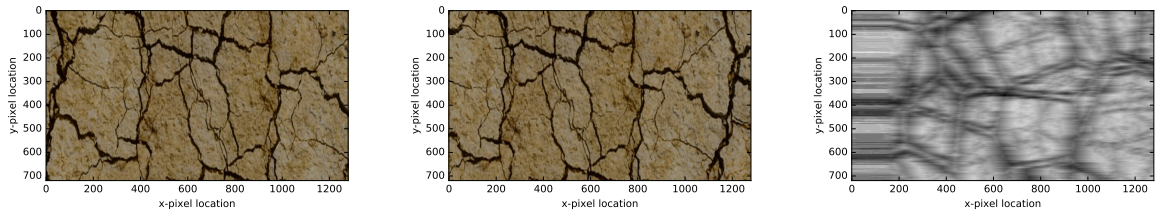
Finally, the 3D point in the right camera coordinate system can be transformed into the world coordinate frame by converting to homogeneous coordinates and applying a coordinate transformation ${}^{w}_{r}\mathbf{H}$

$${}^{r}\mathbf{p} = \begin{bmatrix} X_r \\ Y_r \\ Z \end{bmatrix} \rightarrow {}^{w}\tilde{\mathbf{p}} = {}^{w}_{r}\mathbf{H}\,{}^{r}\tilde{\mathbf{p}}. \tag{7}$$

Determining the 3D coordinates of a point ${}^{w}\mathbf{p}$ therefore requires us to know the camera separation b, the focal length of the camera f, the disparity d of points, and the coordinate transformation ${}^{w}_{r}\mathbf{H}$. It is common for the camera separation and focal length of a stereo system to be known. However, it is more difficult to know the disparity and coordinate transformation.

The disparity may be found using a variety of methods. In general, there are sparse or dense disparity algorithms. Sparse methods track a limited number of features in an image (Szeliski, 2011), whereas dense methods compute disparities for every pixel within an image. A summary of dense disparity methods is provided by Scharstein and Szeliski (2002). The theory we adopt for airborne stereo sensing is applicable regardless of the disparity method chosen. We choose to determine the disparity using a dynamic programming solution for finding shifts in images known as Smooth Dynamic Image Warping (SDW) (Arias, 2016). SDW extends Dynamic Image Warping (DIW) (Hale, 2013) by recovering sub-sample shifts. We choose the SDW algorithm because it provides dense and sub-pixel disparity values for each pixel in an image. We denote the left and right image rows as vectors $\mathbf{l}$ and $\mathbf{r}$ where each element of $\mathbf{l}$, denoted l_i, is approximately an element in $\mathbf{r}$, r_{i+d_i}

$$l_i \approx r_{i+d_i} \;\; for\; i = 1, ..., W. \tag{8}$$



(a) (b) (c)

Figure 2: Left (a) and right (b) corresponding to one frame of the stereo video with the associated disparity image (c). Larger disparities are displayed lighter than small disparities.

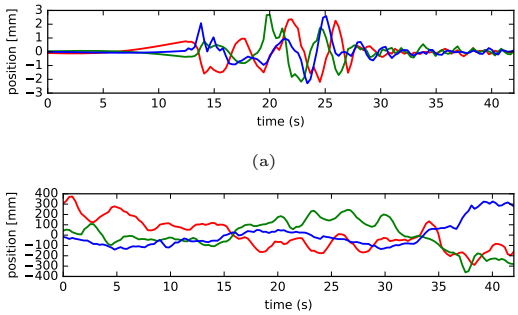
Where W is the size of a row in the image and d_i denotes the disparity found using SDW used in equation 5 to determine the 3D point at pixel i. We estimate the disparity for all pixels in a row, and therefore recover a dense set of 3D point coordinates using SDW. In practice, SDW minimize the absolute error between $\mathbf{l}$ and a shifted version $\mathbf{l}$.

After finding the disparity using SDW, we obtain the 3D location of a point represented in the right camera coordinate frame from equations 5 and 6. The right camera coordinate frame may be translating and rotating relative to the world coordinate frame while the stereo video is being recorded, for example, if the cameras are mounted on a hovering drone. It is necessary to examine the 3D points in a consistent coordinate frame, therefore we require the coordinate transformation ${}^{w}_{r}\mathbf{H}$ for each frame of the video. This coordinate transformation is obtained from information provided by an onboard IMU or GPS.

Many points are observed by the stereo cameras, which are used to form a probability density function (PDF) of ground position at each frame. The PDF is constructed after the coordinate transformation to the world frame, providing consistent measurements with associated uncertainties. In the next section, we use the PDFs to describe earthquake motion recovered from stereo videos.

NUMERICAL EXAMPLES

We test the feasibility of using a drone-mounted stereo vision system using realistic virtual simulations. These simulations provide flexible, precise, and known control of drone and ground motion, allowing us to simulate various drones, cameras, and ground motion signals. We choose to simulate two cameras separated by $b = 30$ cm in the x-direction of the left camera frame. Each camera has a focal length of 4.15 mm, a resolution of 1280×720 pixels, and an image sensor element size of 3.75 microns. During the simulation, the cameras view a moving ground surface from a height of about 2 m. The simulated vibrating ground surface sports various cracks and a texture representing cracked clay (see Figure 2a). An earthquake signal (Figure 3a) recorded at USGS station OK034 in fall of 2016 near Cushing, Oklahoma is used to displace the surface in three dimensions (Center for Engineering Strong Motion Data, 2016). We show results for the recovered displacement of the ground surface with and without drone motion during the earthquake, respectively. We first show earthquake motion recovery when the drone is not moving. Videos are rendered for both the left and right cameras as the ground moves in three dimensions according to the ground position curves in Figure 3a. The videos are used to deduce a PDF representing ground motion. We clarify that when estimating disparities using the SDW algorithm, we use a sub-sample precision of $1:50$ and a strain range of ± 0.02. The sub-sample precision we choose is sufficiently small to recover subtle shifts between left and right images; a finer sub-sample precision would not degrade results, but would be more computationally expensive. The small and narrow strain range we choose implies that the shifts we seek are smoothly varying along the image rows. We display the point



(a)

(b)

Figure 3: (a) three-component USGS station OK034 earthquake signal and (b) drone motion (with mean position subtracted) used in simulations. The colors red, blue, and green correspond to x, y, and z, respectively.

location PDF in the world coordinate frame for each frame in the video, as shown in Figure 4a. The PDF is characteristically smooth, broad, and sports horizontal bands. The horizontal banding in the recovered PDF is a result of the discrete disparity values recovered using SDW. We observe in Figure 4c that the mean of the PDF, after subtracting the initial distance to the ground, traces the true displacement of the ground sur-

face. The true and observed displacement curves are on top of one another, the difference between the true and observed signals (Figure 4c) shows the signal is recovered well. The error standard deviation is 0.011 mm, which is five orders of magnitude less than the camera height (h=2 m).

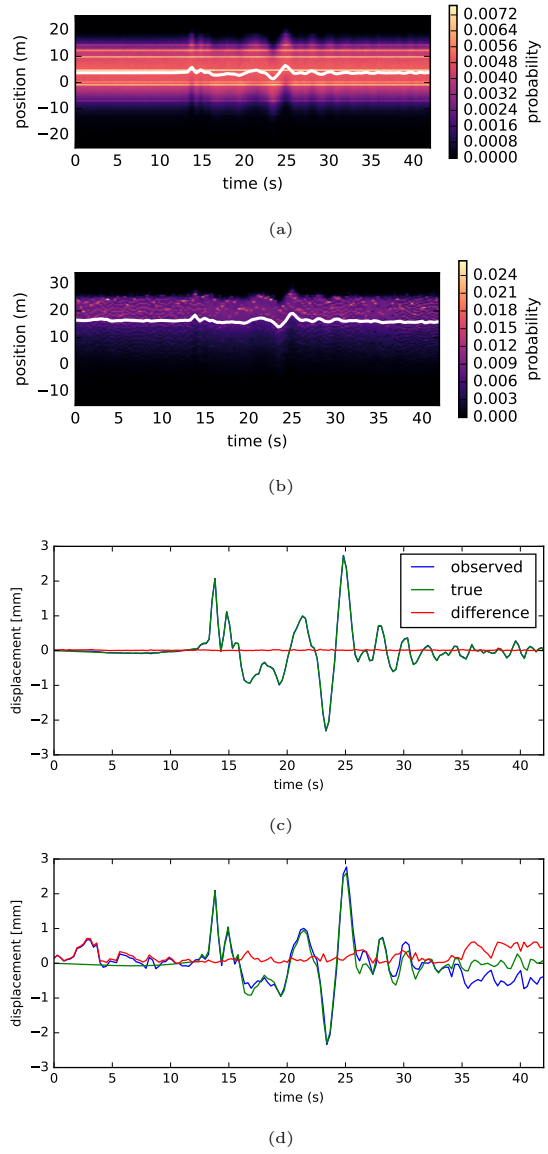


Figure 4: (a)-(b) Recovered PDFs for experiments 1 and 2; (c)-(d) observed signals compared to true. The larger signal amplitudes are well recovered when drone motion is present.

We repeat the first experiment (Figure 4a) with known 3D drone motion. We simulate drone motion according to real-life motion measured from a DJI Matrice 100 drone. We gather the real-life motion by instructing the drone to hover at a fixed position, the drone motion is shown in Figure 3b. The drone moves several orders of magnitude more than the earthquake signal we wish to recover. We display the PDF of point location in world coordinates for the second experiment in Figure 4b. The banding observed in the PDF for experiment 1 is not present in the PDF for experiment 2, because of the coordinate transformation we use to correct for drone motion. However, we see their remnants as scattered peaks in the recovered PDF which seem to mimic the vertical motion of the drone.

We expect the error to deteriorate when disparities are small, when the drone is farther away from the ground. Thus, the displacement error we obtain when the drone is moving depends on the drone motion. The observed versus true ground motion for experiement 2 is shown in Figure 4d. We see that the large amplitudes of the earthquake signal are well recovered, and that our error is orders of magnitude smaller than the drone motion. However, the errors present when the drone is moving are larger than when the drone is stationary. We observe the that the difference for experiment 2 sports a standard deviation of 0.30 mm. Based on our results, we conclude that a stereo camera is a viable option for monitoring earthquake-like sub-millimeter ground motion. The simulation parameters we use reflect realistic drone and camera parameters.

CONCLUSIONS

The drone motion, earthquake motion, and camera specifications used in this work are derived from data gathered in the real world. We assume known relative drone position during acquisition based on IMUs and GPS. A dense disparity algorithm is used to deduce many estimates of ground position at a given time, which are combined to form a statistical representation of ground position. Sub-pixel disparity estimates are necessary to measure the small ground motions. The strategy for disparity estimates used here does not allow us to recover lateral ground displacement; future work focuses on augmenting our current method to recover lateral ground displacement. We conclude that drone mounted stereo cameras have the potential to measure seismic signals, with the added benefit of sensor mobility and measurement uncertainty.

ACKNOWLEDGEMENTS

We thank the sponsor companies of the Center for Wave Phenomena, whose support made this research possible. We also acknowledge using the open source computer graphics engine Blender for our videos, and appreciate those who contributed to its development.

REFERENCES

Arias, E., 2016, Seismic slope estimation: Analyses in 2D/3D: PhD thesis, Colorado School of Mines. Arthur Lakes Library.

Bostick III, F., 2000, Field experimental results of three-component fiber-optic seismic sensors, *in* SEG Technical Program Expanded Abstracts 2000: Society of Exploration Geophysicists, 21–24.

Center for Engineering Strong Motion Data, 2016, USGS Station OK034: www.strongmotioncenter.org.

Daley, T. M., B. M. Freifeld, J. Ajo-Franklin, S. Dou, R. Pevzner, V. Shulakova, S. Kashikar, D. E. Miller, J. Goetz, J. Henninges, et al., 2013, Field testing of fiber-optic distributed acoustic sensing (das) for subsurface seismic monitoring: The Leading Edge, **32**, 699–706.

Hale, D., 2013, Dynamic warping of seismic images: Geophysics, **78**, S105–S115.

Mateeva, A., J. Lopez, J. Mestayer, P. Wills, B. Cox, D. Kiyashchenko, Z. Yang, W. Berlang, R. Detomo, and S. Grandi, 2013, Distributed acoustic sensing for reservoir monitoring with vsp: The Leading Edge, **32**, 1278–1283.

Mateeva, A., J. Lopez, H. Potters, J. Mestayer, B. Cox, D. Kiyashchenko, P. Wills, S. Grandi, K. Hornman, B. Kuvshinov, et al., 2014, Distributed acoustic sensing for reservoir monitoring with vertical seismic profiling: Geophysical Prospecting, **62**, 679–692.

Mateeva, A., J. Mestayer, B. Cox, D. Kiyashchenko, P. Wills, J. Lopez, S. Grandi, K. Hornman, P. Lumens, A. Franzen, et al., 2012, Advances in distributed acoustic sensing (das) for vsp, *in* SEG Technical Program Expanded Abstracts 2012: Society of Exploration Geophysicists, 1–5.

Mestayer, J., B. Cox, P. Wills, D. Kiyashchenko, J. Lopez, M. Costello, S. Bourne, G. Ugueto, R. Lupton, G. Solano, et al., 2011, Field trials of distributed acoustic sensing for geophysical monitoring, *in* SEG Technical Program Expanded Abstracts 2011: Society of Exploration Geophysicists, 4253–4257.

Quam, L. H., R. B. Tucker, S. Liebes Jr, M. J. Hannah, and B. G. Eross, 1972, Computer interactive picture processing: Technical Report 166, Stanford Artificial Intelligence Project Memo.

Scharstein, D., and R. Szeliski, 2002, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms: International journal of computer vision, **47**, 7–42.

Szeliski, R., 2011, Computer vision: Algorithms and applications: Springer.