Airborne seismic monitoring using stereo vision

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

ABSTRACT

Environmental impact and the high cost of acquiring land seismic data are major factors to consider when designing a seismic survey. We explore means to quickly record seismic data without disturbing, or contacting, the ground surface, while reducing environmental impact and cost. Recent developments in computer vision techniques and drone technology lead us to propose passively observing ground displacement with a drone-borne stereo video camera system. The recovered displacement is represented as a time varying probability distribution function (PDF) of ground displacement. Using this PDF, we can estimate the uncertainty of the displacement measurements. We conclude that currently available camera and drone systems may be used to measure sub-millimeter ground displacements, with associated uncertainties.

Key words: airborne, contactless seismic, stereo vision, acquisition

1 INTRODUCTION

Geophysical survey design requires one to consider the environment, access, and acquisition cost. When operating in environmentally sensitive areas, it may not be possible to conduct a ground-based survey without disturbing an ecosystem. In addition, in remote locations a survey may not be feasible due to access or other logistical concerns. Financial limitations also influence data acquisition feasibility. For example, when performing a seismic survey, a large amount of equipment, time, and personnel are required, leading to high acquisition cost. The cost may be reduced by optimizing survey design parameters such as seismic source geometry, source signal length and frequency content, and receiver spacing. Technological advancements can also reduce costs and potentially lead to new opportunities for exploration; an excellent example is distributed acoustic systems (DAS) (Bostick III, 2000; Mestayer et al., 2011; Mateeva et al., 2012; Daley et al., 2013; Mateeva et al., 2013, 2014). We advocate in this paper, that recent developments in computer vision and robotics also have the impact potential for seismic data acquisition, leading towards fast, low cost, and environmentally friendly airborne seismic surveying.

Monitoring motion without contacting a surface can be performed using passive or active systems. Laser Doppler vibrometers (LDV) are examples of active systems, which use known source signals and the phase of the return signals to deduce the distance the signal travelled to a target. LDV's have been used in the past to investigate the feasibility of remotely detecting ground motion from seismic waves (Berni, 1991, 1994). Other active systems uses microwaves, and interferometry, to deduce vibrations and resonant frequencies in a structural engineering context such as in Stanbridge et al. (2000). In contrast to active systems, passive systems do not require a source for deducing motion and leverage high speed video cameras and ambient lighting. Here we focus on passive methods for deducing motion.

We advocate measuring ground displacement as a function of time using stereo vision theory. As our left and right eyes allow for depth perception, two images taken from laterally offset cameras allow for a distance estimate. Stereo vision has been used in the past to passively deduce distance using two images taken from laterally offset cameras (Quam et al., 1972). We measure the lateral shift, or disparity, between points observed in the two images. As we show later, this disparity is inversely proportional to the distance from the camera to the points observed in the images. Therefore, the process of finding ground displacement from stereo images requires us to accurately compute perceived shifts between two images. The stereo vision process is discussed in more detail in the Theory section.

A drawback of stereo vision is that small disparity errors result in large distance errors; we therefore must be able to precisely measure shifts between images in order to obtain a reliable distance measurement. To improve our ability to measure small shifts between images, we can amplify the shifts using a recent and exciting advancement in computer vision called motion magnification. Motion magnification allows us to remotely monitor visually imperceivable vibrations using passive high speed cameras (Wahbeh et al., 2003; Wadhwa et al., 2014; Rubinstein et al., 2014). The technique has been used by many others to solve various scientific engineering problems.

Chen et al. (2014) use the method for deducing structural information of a steel beam using video; a similar approach is taken by Shariati et al. (2015) to deduce natural structural vibration modes of other structures. Remotely monitoring heartbeats using video is another intriguing use of motion magnification (Wu et al., 2012). Davis et al. (2014) recover sound from video alone by filming light objects, e.g. a bag of chips, subject to tiny vibrations caused by someone speaking in the vicinity. It is our intention to apply motion magnification, in conjunction with stereo vision, in order to acquire seismic data without contacting the ground. Although do not include motion magnification in this paper, we note that contactless motion can benefit significantly from this new development.

In this paper, we demonstrate and detail a method for sampling a seismic wavefield, represented by ground displacements, that has potential to reduce costs while providing new opportunities for geophysical exploration or earthquake monitoring. First, we clarify the stereo vision theory used to determine ground displacement variations with time. Using the collection of ground displacement values, we construct a probability density function (PDF) of ground displacement, from which we deduce the uncertainty of our measurement. Lastly, we conduct a realistic computer graphics simulation of a real-life earthquake signal, demonstrating the feasibility of using stereo videos to recover sub-millimeter displacement signals, with associated uncertainties, from a moving airborne platform.

2 THEORY

We aim to passively measure ground displacement, without touching the ground, and obtain an uncertainty estimate of our measurement. In this section, we provide a brief summary of how stereo vision to acquire a ground displacement measurement with associate uncertainty. The stereo vision theory we summarize is common in standard computer vision literature (Szeliski, 2011). Stereo videos are represented as a sequence of frames taken from two offset cameras (left and right), as shown in Figure 1. This geometry allows one to acquire the distance to points viewed by the stereo cameras using geometric relations based on similar triangles and triangulation.

We begin by defining the origin of three coordinate frames representing the origins of the world, left camera, and right camera. It is our goal to consistently represent a PDF of ground position in the world coordinate frame using images from stereo cameras. 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. We adopt the notation where preceding superscripts denote the coordinate frame of a point and subscripts denote the label of a specific point as w, l, and r for the world, left camera, and right camera coordinates, respectively. The z-axis of the world coordinate frame points down, a left camera is placed at a height h above the world origin. The right camera origin is placed a distance b away from the left camera in the x-direction and is noted by ${}^{l}\mathbf{o}_{r} = [b,0,0]^{T}$. We represent transformations between an arbitrary coordinate system a to another coordinate system b using a 4×4 matrix ${}^{b}_{a}\mathbf{H}$. This matrix may be obtained by augmenting the rotation matrix from coordinate system a to b, denoted ${}^{b}_{a}\mathbf{R}$, with the origin of a in b, denoted ${}^{b}\mathbf{t}_{a,org}$

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

We use a homogeneous point representation in order to apply rotation and translation of a point using a single matrix-vector multiplication. A point $\mathbf{x}$ in 3D space may be represented in homogeneous coordinates as a four-element point $\tilde{\mathbf{x}}$ by appending a fourth arbitrary element to $\mathbf{x}$. We represent a point $\mathbf{x}$ in homogeneous coordinates using the notation $\tilde{\mathbf{x}}$, and note that homogeneous coordinates that are scalar versions of one another are considered equivalent points in 3D space. We may obtain the original coordinate from the homogeneous coordinate by dividing by the fourth element:

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

With homogeneous coordinates, we can transform the right camera origin from the left camera frame to the world frame by

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

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

$$ {}^{w}\tilde{\mathbf{x}} = {}^{w}_{r}\mathbf{H}\,{}^{r}\tilde{\mathbf{x}}. \tag{4} $$

Stereo vision allows us to determine the 3D location of a point $\mathbf{p}$ observed using the left and right cameras. We can represent the point 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 right camera frame. Using stereo vision, we find ${}^{r}\tilde{\mathbf{p}}$ then transform the point to the world coordinate frame 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. We denote these pixel coordinates as (x_l, y_l) and (x_r, y_r) for the left and right camera, respectively. Projecting a 3D point, such as ${}^{r}\mathbf{p}$, onto a 2D image can be described using a pinhole camera model, which assumes that all 3D points project onto the image plane of a camera through a single point, or pinhole. The model is appropriate to use when a camera has been calibrated to remove lens distortion, which is commonly performed on images and (we assume that all images used in this paper have already been calibrated). 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 left and right x-pixel coordinates of the point as a function of the camera

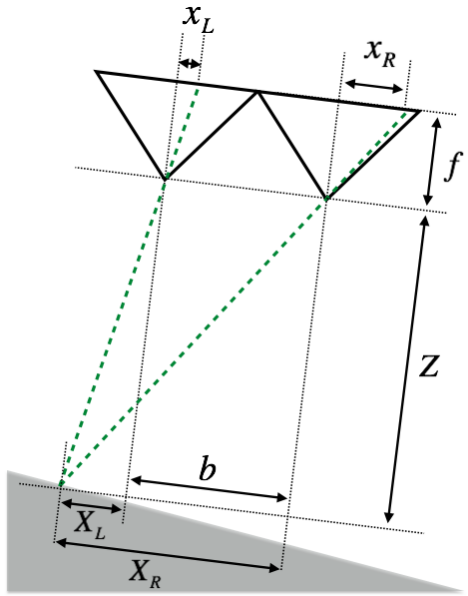


Figure 1. A schematic of stereo cameras viewing the ground from above. The two cameras have optical axes that are aligned but separated by a distance b. Both cameras have a focal length of f. The dashed paths show where a point on the ground projects on the left and right cameras. Note that the cameras do not have to be aligned with the ground surface, however the recovered depth Z is measured from the cameras orthogonal to the lines connecting the pinholes.

focal length f, in pixels, and the distance to the point Z

$$\begin{aligned} \frac{x_l}{f} &= \frac{X_l}{Z} \rightarrow x_l = \frac{fX_l}{Z} \\ \frac{x_r}{f} &= \frac{X_r}{Z} \rightarrow x_r = \frac{fX_r}{Z}. \end{aligned} \tag{5}$$

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

$$d = x_r - x_l = \frac{fX_r - f(X_r - b)}{Z} = \frac{fb}{Z}. \tag{6}$$

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 ${}^{r}\tilde{\mathbf{p}}$

$$\begin{aligned} X_r &= \frac{Zx_r}{f} \\ Y_r &= \frac{Zy_r}{f}. \end{aligned} \tag{7}$$

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 ${}^{w}_{r}\mathbf{H}$

$${}^{r}\mathbf{p} = \begin{bmatrix} X_r \\ Y_r \\ Z \end{bmatrix} \rightarrow {}^{w}\tilde{\mathbf{p}} = {}^{w}_{r}\mathbf{H}\,{}^{r}\tilde{\mathbf{p}}. \tag{8}$$

The pseudocode in Algorithm 1 outlines how stereo vision is used for recovering ground motion.

Algorithm 1 Computing point location PDF from stereo images

```
Let f denote the focal length of the camera (in pixels)
Let b denote the left and right camera separation
Let superscript r denote right camera coordinate frame
Let superscript w denote world camera coordinate frame
Let ^w_rH denote the coordinate transformation from the right camera to world coordinates
for each frame in frames do
  for each pixel in frame do
    Let (x, y) denote the pixel coordinates in this frame
    Let ^rp = [X, Y, Z]^T denote the recovered 3D point coordinates on the ground
    Compute pixel disparity d at (x, y)
    Compute → Z = fb/d → X = Zx/f → Y = Zy/f → ^rp
    Compute point location ^wp = ^w_rH ^rp
  end for
  Compute ground location histogram using all ^wp in this frame
  Compute ground location PDF from the histogram for this frame
end for
```

Determining the 3D world 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 than b and f.

The disparity may be found using a variety of methods. In general, there are two main categories of disparity algorithms that use sparse or dense point correspondence. Sparse methods track a limited number of features in an image and have been in development since the 1970's (Szeliski, 2011), whereas dense methods compute disparities for every pixel within an image (Scharstein and Szeliski, 2002). The theory we adopt here 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. As a dynamic programming approach to finding shifts between two images, SDW works by minimizing a constrained nonlinear optimization problem. In summary, SDW determines shift values between two images that may be used to 'warp' one image to the other. In our context, we determine shifts between the left and right stereo images. Given that the cameras are only separated in the x-direction, the shift between images will only be in the x-direction. Therefore, we estimate 1D shifts between the rows of left and right images. We note that estimating 2D shifts provides more constraints to the optimization problem and may be beneficial; however, we use 1D shifts in this paper for simplicity and computational efficiency. We can denote the 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}

$$l_i \approx r_{i+d_i} \textit{ for } i = 1, ..., W, \tag{9}$$

where W is the width of the image in pixels and d_i denotes the disparity used in equation 6 to determine the 3D point which projects onto 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, we minimize the error between $\mathbf{l}$ and a shifted version $\mathbf{l}$; SDW minimizes absolute error, which results in shifts that express how each element of $\mathbf{r}$ may be shifted to in order for a match $\mathbf{l}$. The details of how the SDW minimization problem is implemented are beyond the scope of this paper but can be found in (Hale, 2013) and (Arias, 2016). After finding the disparity using SDW, we obtain the 3D location of a point represented in the right camera coordinate frame from equations 6 and 7.

The cameras 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 may be provided from an onboard IMU, which uses accelerometers and gyroscopes to monitor position and orientation. In addition, other sensors, such as GPS may be used to monitor the camera motion during flight.

Many points on the ground are simultaneously viewed by the stereo cameras, and therefore we may statistically represent the position of the ground viewed by the system as a Probability Density Function (PDF). This representation of the ground assumes that between frames, all points undergo the same translation. A PDF of ground position is therefore available for each frame in the stereo video, which provides quantitative uncertainty information of the ground position. The PDFs change between frames as the ground and cameras move, however we may isolate the ground and camera motion by measuring and removing the camera motion. To remove the camera motion, we perform a coordinate transformation from right camera to world coordinate frame using equation 1 to form ${}^{w}_{r}\mathbf{H}$ for a particular frame. We note that the transformation matrix ${}^{w}_{r}\mathbf{H}$ contains camera motion information that is independent of the ground motion e.g. from an IMU. After camera motion is removed, we obtain a PDF representing the position of the ground in a consistent coordinate system as it varies with time. Summary statistics, such as the mean, may be computed from the PDF to provide us with a 'trace' that represents ground motion. Similarly, we may compute the variance of the PDF as a proxy for the uncertainty of our measurements. Therefore measurements of the ground position, with associated uncertainty, are available from images of the ground taken remotely.

We can analytically determine how error in projected pixel locations translates into disparity error. We may quantify the change in disparity given a small change, or error, in the pixel locations Δx_r or Δx_l using total derivatives, which for a function $z = f(x, y)$ can be approximated as

$$\Delta z \approx \frac{\delta f}{\delta x}\Delta x + \frac{\delta f}{\delta y}\Delta y, \tag{10}$$

where small changes in x and y are denoted as Δx and Δy, respectively. Considering the disparity, $d(x_r, x_l)$, from equation 6 we can compute the error

$$\begin{aligned} \Delta d &= \frac{\delta d}{\delta x_r}\Delta x_r + \frac{\delta d}{\delta x_l}\Delta x_l \\ &= \Delta x_r - \Delta x_l, \end{aligned} \tag{11}$$

where Δx_r and Δx_l are errors in the point projection on the right and left camera, respectively. We expect the errors in the left and right cameras to be independent, as they were taken from independent cameras, and assume the errors are zero mean. The expected value of the disparity error then becomes:

$$\mu_d = E[\Delta d] = E[\Delta x_r - \Delta x_l] = E[\Delta x_r] - E[\Delta x_l] = 0. \tag{12}$$

The variance of the disparity error can be computed by leveraging equation 12 and using the pixel error independence assumption to obtain

$$\sigma_d^2 = E[(\Delta d - \mu_d)^2] = \sigma_L^2 + \sigma_R^2. \tag{13}$$

We conclude that the mean of the disparity error to be zero given independent and zero mean errors in the left and right camera pixel projection coordinates. The variance of the disparity is then the sum of the variance of the errors in the left and right cameras.

The depth estimate error given an error in disparity de-

fined in equation 6 is

$$\Delta Z = \frac{\delta Z}{\delta d}\Delta d = -\frac{fb}{d^2}\Delta d. \tag{14}$$

We see that the absolute value of the distance error is proportional to $\frac{fb}{d^2}$. Proportionally, this result implies that an increase in camera separation (b) or focal length (f) increases the distance error for a given disparity (d). However, the disparity dominates the depth error as $\frac{1}{d^2}$, implying that large disparities can drastically decrease the depth error. The mean of the error in depth Z can be found by leveraging equation 14 and simplifying the ΔZ term,

$$\mu_Z = E[\Delta Z] = -\frac{fb}{d^2}E[\Delta d] = 0, \tag{15}$$

while the variance of the error in depth Z is,

$$\sigma_Z^2 = E[(\Delta Z - \mu_Z)^2] = \left(\frac{fb}{d^2}\right)^2 \sigma_d^2. \tag{16}$$

The standard deviation in the depth error is proportional to the quantity $\left(\frac{fb}{d^2}\right)^2$. This quantity is dominated by the effects of d^{-4}; if the disparity increases, then the variance in the depth error rapidly decreases. While mean and variance summarize the behavior of the distance error, a more complete representation of the displacement is represented using PDF's, as demonstrated in the next section.

3 NUMERICAL EXAMPLES

We asses the feasibility of measuring ground motion with a drone-mounted stereo vision system using realistic virtual simulations. Virtual simulations provide flexibility and precise control of drone and ground motion, allowing us to simulate various drones, cameras, and ground motion signals. By using simulations, we know true ground and camera motion, can be compared to the recovered motion using stereo vision. We 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. These camera parameters are feasible and common, and were chosen to reflect current smartphone cameras. The cameras view a simulated moving ground surface from a height of 2 m, a reasonable height for a drone hovering above the ground.

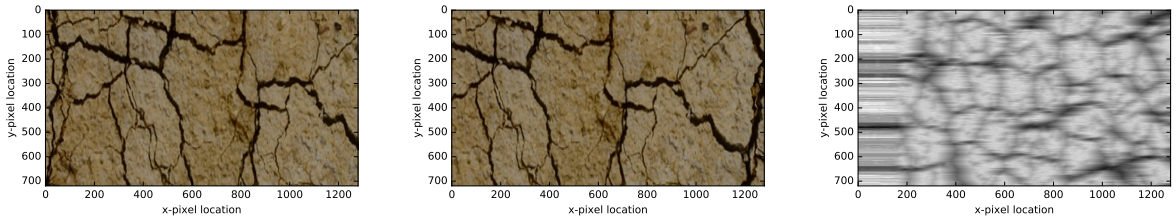
The vibrating ground surface is described by texture and topography representing cracked clay (Figure 2). We note that the appearance of cracks in the images intentionally does *not* reflect the true cracks in the ground surface. The true cracks on the ground surface are characterized by Vorinoi noise and have a fractal-like pattern, as shown in the recovered disparity image (Figure 2(c)). We intentionally separate the appearance of the ground from its true topography to show that the recovered disparities are influenced by the topography, rather appearance.

An earthquake signal (Figure 3) recorded at USGS station OK034 in the fall of 2016 near Cushing, Oklahoma is used to displace the surface in three dimensions (Center for Engineering Strong Motion Data, 2016). We note that the displacement of the signal is on the order of 3 mm, with many displacements below 1 mm. The vibrating ground surface is observed using the virtual stereo cameras, which hover above the ground (see Figure 4) as if mounted on a drone. In the following, we present recovered ground displacement results with and without drone motion.

We first show displacement recovery when the drone is stationary. Videos are rendered for both the left and right cameras as the ground moves in three dimensions according to the ground position curves shown in Figure 3. Dense disparity values are computed at each time step, which are used to construct a PDF representing the ground location as viewed from the stereo cameras (see Algorithm 1. When estimating disparities using the SDW algorithm, we use a sub-sample precision of $1 : 50$ and a strain range of ± 0.02. The strain range chosen is low and narrow because of prior knowledge we posses: the disparities we seek are smoothly varying along a row. Setting a wider strain range allows for more rapidly varying shifts along a row, and may be more appropriate when viewing a rugged surface. Here, the simulated ground is relatively flat, thus implying that the observed shifts vary smoothly.

The sub-sample precision chosen is sufficiently small to recover minute shifts between left and right images. We note that a higher sub-sample shift precision may be used, at the expense of an increase in computational memory requirements, however one may not observe benefits beyond a certain sub-sample shift precision value (Arias, 2016). In contrast, choosing too low of a sub-sample shift precision value may lead to poor shift recovery. We test larger sub-sample precision values and found little benefit beyond $\frac{1}{50}$. The sub-sample shift value is best estimated by performing disparity computations on the first frame of the video and examining the recovered shifts. If the recovered shifts vary wildly, or seem discontinuous, then a higher degree of sub-sample precision may be required. We note that if computational time and memory is not an issue, then one may set the value of sub-sample shift very high without risking the accuracy of recovered shifts. An example of recovered disparities for the first frame is shown in Figure 2(c). Note that the recovered disparities are somewhat smooth along a row, and they resemble cracks which reflect the cracks in the simulated ground surface.

Figure 5(a) shows the recovered point location PDF in world coordinate frame for each frame in the video. The PDF gives insight on the statistical characteristics of our measurement. The PDF is smooth, broad, and contains horizontal bands. The horizontal banding in the recovered PDF is present due to discrete disparity values recovered using SDW, which are of limited precision. Regardless, we observe in Figure 6(a) that the mean of the PDF, after subtracting the initial distance to the ground, traces the true displacement of the ground surface. The true and observed displacement curves are on top of one another, which shows that the signal has been recovered well. The error, for this example, has a standard deviation of 0.011 mm. The standard deviation of the error is within a hundredth of a millimeter; five orders of magnitude less than the height at which the measurement was taken (h=2 m).



(a) (b) (c)

Figure 2. Appearance of (a) left and (b) right images recovered for one frame and (c) associated disparity image. The images resemble a cracked clay surface, and are laterally shifted versions of one another due to the stereo camera setup. The topography of the simulated ground surface is intentionally different from the image texture, showing that ground topography dominates the recovered disparity appearance.

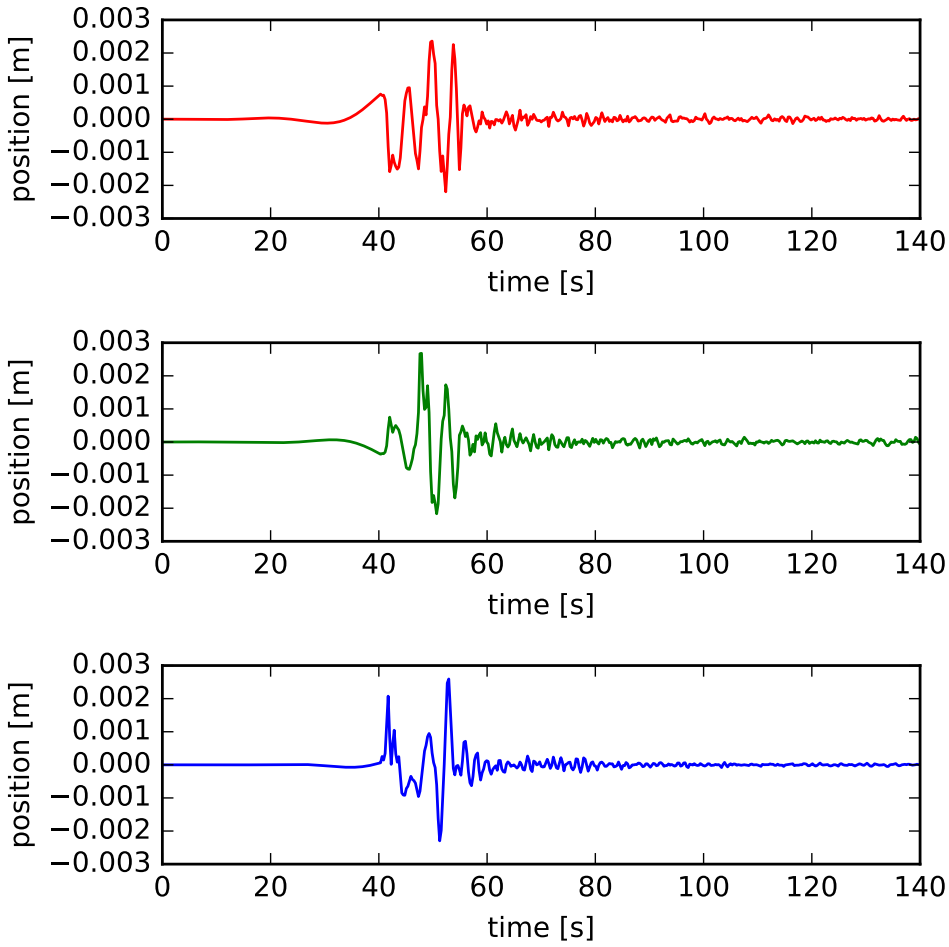


Figure 3. Earthquake signal recorded in the fall of 2016 by USGS station OK034 near Cushing, Oklahoma. The red, blue, and green curves show the ground displacement in x, y, and z directions.

We repeat the experiment (Figure 5(a)) with a moving drone and the same ground motion signal. When the drone is moving, the range of disparity values we must search through using SDW is considerably larger, making disparity estimates more computationally costly. We simulate drone motion which resembles real-life motion we measured from a DJI Matrice 100 drone, instructed to hover at a fixed position for about two minutes. The time spanned by the earthquake signal in Figure 3 is roughly two minutes, as shown in Figure 4. The drone moves a considerable amount during the simulation, roughly a

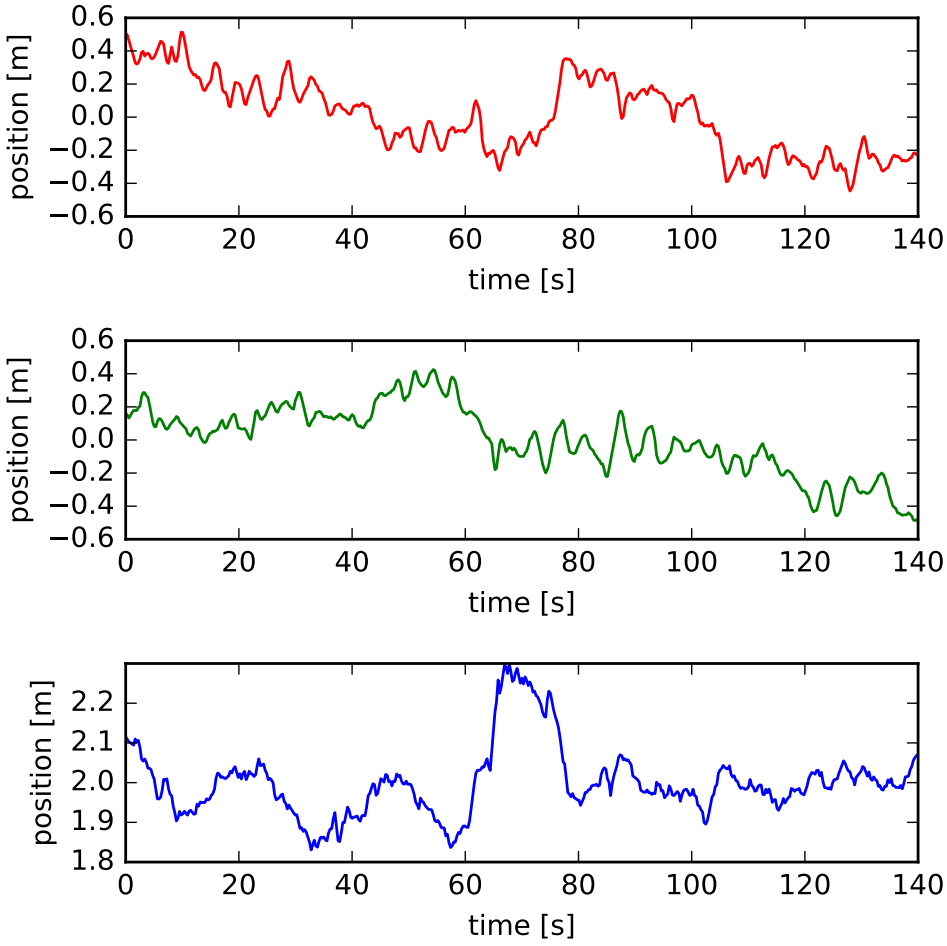


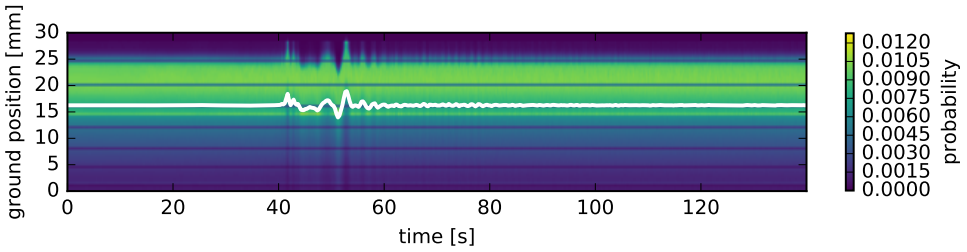
Figure 4. Recorded drone motion used in hovering drone simulation. We note that the drone motion is three orders of magnitude larger than the signal we wish to recover in Figure 3.

quarter of a meter; this variation is orders of magnitude larger than the earthquake signal we wish to recover. The recovered PDF of point location in world coordinates for this second experiment is shown in Figure 5(b). We note that the banding observed in the PDF for the first experiment is not present in the PDF for second experiment 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. The errors present when the drone is moving are larger than when the drone is stationary, we observe an error standard deviation of 0.30 mm.

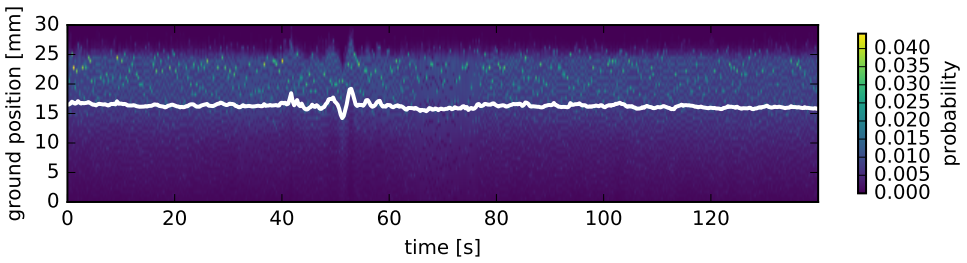
We expect the error to deteriorate when disparities are small, i.e. when the drone is farther away from the ground. Thus, the displacement error we obtain when the drone is moving depends on the drone motion. We confirm this expectation by showing the drone height and absolute error for vertical ground motion in Figure 7. As the drone approaches the ground, the effective area represented in a pixel increases and the disparity estimates are larger. As previously described in the Theory section, we expect the error in vertical position (equation 16) to drastically improve for larger disparities. Based on our results, we conclude that a stereo camera can feasibly measure earthquake-like sub-millimeter ground motion. The simulation parameters we use reflect realistic drone and camera parameters.

4 CONCLUSIONS

The drone motion, earthquake motion, and camera specifications used in this work reflect real-life data. We assume known drone position during acquisition, which may be available from IMUs and GPS. A dense disparity algorithm leads to many estimates of ground position at a given time, in turn allowing access to a statistical representation of ground position. We note that sub-pixel disparity estimates, provided by the disparity algorithm, are necessary to characterize the sub-



(a)



(b)

Figure 5. Recovered ground position PDF with (a) no drone motion and (b) realistic drone motion. White curves show the mean of each PDF, which resemble the vertical ground motion from the simulated earthquake.

tle earthquake ground motions. The strategy for disparity estimates used here dose not provide point correspondence in time, and therefore does not allow us to recover lateral ground displacement. Future and ongoing work focuses on augmenting our current method to recover lateral ground displacement. We conclude that drone-borne stereo cameras can potentially be used to measure seismic signals, with associated uncertainties.

5 ACKNOWLEDGEMENTS

We thank the sponsor companies of the Center for Wave Phenomena, whose support made this research possible. We thank Matt Bailey and Ross Bunker for help acquiring drone position data. In addition, we thank Dr. David Wald for his feedback on the presented work. Lastly, we acknowledge using the open source computer graphics engine Blender for our videos, and appreciate those who contributed to its creation.

REFERENCES

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

Berni, A. J., 1991, Remote seismic sensing. (US Patent 5,070,483).

———, 1994, Remote sensing of seismic vibrations by laser Doppler interferometry: Geophysics, **59**, 1856–1867.

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.

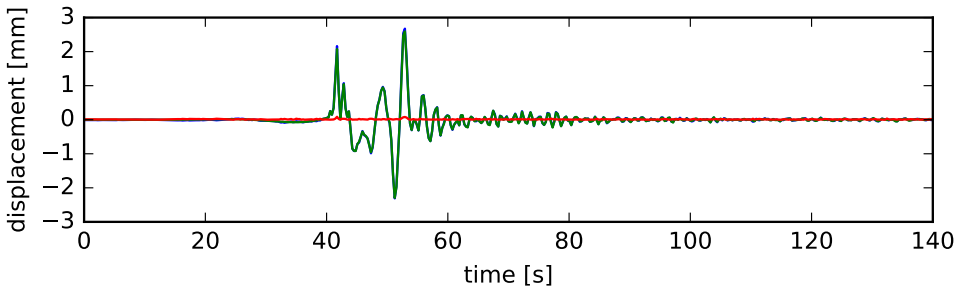
Chen, J. G., N. Wadhwa, Y.-J. Cha, F. Durand, W. T. Freeman, and O. Buyukozturk, 2014, Structural modal identification through high speed camera video: Motion magnification, *in* Topics in Modal Analysis I: Springer, **7**, 191–197.

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.

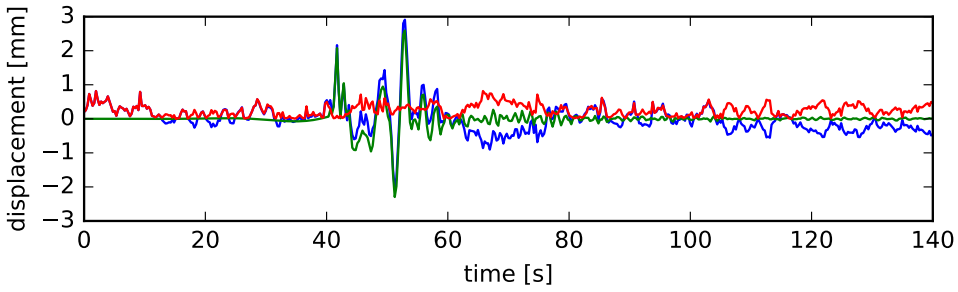
Davis, A., M. Rubinstein, N. Wadhwa, G. J. Mysore, F. Durand, and W. T. Freeman, 2014, The visual microphone: Passive recovery of sound from video: ACM Transaction on Graphics, **33 (4)**, 79:1–79:10.

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

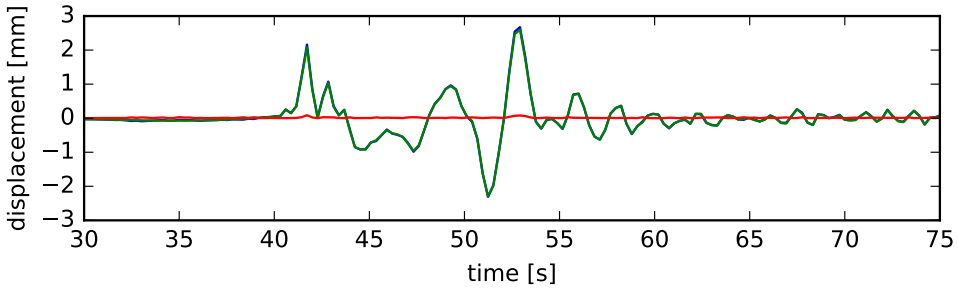
Mateeva, A., J. Lopez, J. Mestayer, P. Wills, B. Cox, D.



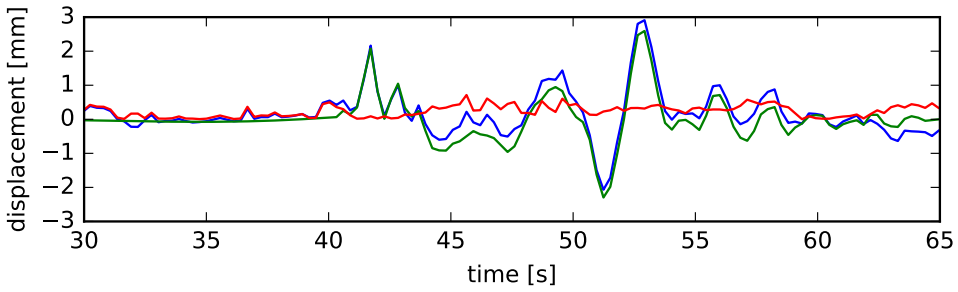
(a)



(b)



(c)



(d)

Figure 6. Recovered ground displacement signal (blue) compared with true displacement (green) (a) without and (b) with drone motion. Red curves show the absolute difference between recovered and true ground motion. Zoomed plots for (a) and (b) are shown in (c) and (d), respectively. Drone motion influences the recovery of the input signal.

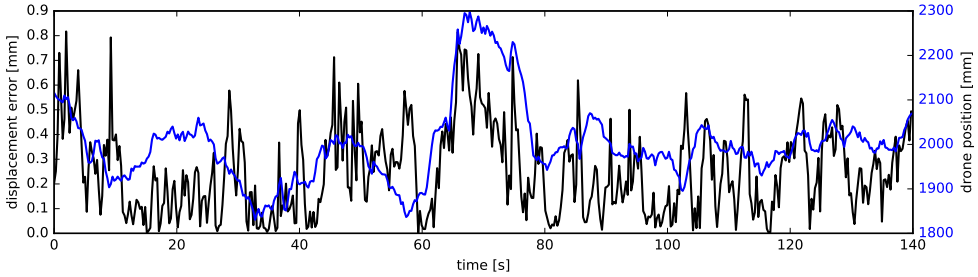


Figure 7. Relationship between signal error and drone height. The error is larger when the drone is farther away from the ground, which explains the errors in the recovered signals in Figure 6(d).

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.

Rubinstein, M., et al., 2014, Analysis and visualization of temporal variations in video: PhD thesis, Massachusetts Institute of Technology.

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.

Shariati, A., T. Schumacher, and N. Ramanna, 2015, Eulerian-based virtual visual sensors to detect natural frequencies of structures: Journal of Civil Structural Health Monitoring, 1–12.

Stanbridge, A., D. Ewins, and A. Khan, 2000, Modal testing using impact excitation and a scanning ldv: Shock and Vibration, **7**, 91–100.

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

Wadhwa, N., M. Rubinstein, F. Durand, and W. T. Freeman, 2014, Riesz pyramids for fast phase-based video magnification: Computational Photography (ICCP), 2014 IEEE International Conference on, IEEE, 1–10.

Wahbeh, A. M., J. P. Caffrey, and S. F. Masri, 2003, A vision-based approach for the direct measurement of displacements in vibrating systems: Smart Materials and Structures, **12**, 785.

Wu, H.-Y., M. Rubinstein, E. Shih, J. V. Guttag, F. Durand, and W. T. Freeman, 2012, Eulerian video magnification for revealing subtle changes in the world.: ACM Trans. Graph., **31**, 65.