Sensor signal processing

Towards Optical Flow in Time-of-Flight Range Imaging Using the Continuous Wavelet Transform

Jordan J. Wise[123], and Lee Streeter[1*]

[1] School of Engineering, University of Waikato, Hamilton, New Zealand
[2] Department of Physics, University of Auckland, Auckland, New Zealand
[3] Dodd-Walls Centre for Photonic and Quantum Technologies, New Zealand
[*] Member, IEEE

Manuscript received XXXX XX, 2023; revised XXXX XX, XXXX; accepted XXXX XX, XXXX. Date of publication XXXX XX, XXXX; date of current version XXXX XX, XXXX.

Abstract—Time-of-flight range imaging cameras perform full-field distance measurement, but require multiple frames to be captured sequentially in order to construct a single range image. When motion occurs during data acquisition, artifacts appear which limit the accuracy of range estimates. Analysis of these artifacts finds that they take the form of spatial waves whose frequency encodes motion information. We therefore propose a new technique for measuring motion in time-of-flight range imaging based on the continuous wavelet transform. Synchrosqueezing is explored as a means of improving the accuracy of velocity estimates.

Index Terms—Time-of-flight range imaging, transverse motion, optical flow, wavelet transform.

I. INTRODUCTION

Time-of-flight (ToF) range imaging is a technique which allows for computationally efficient estimation of depth, measuring distance as a function of the time it takes light to travel from a source to a scene and back to a camera [1]. In indirect ToF, amplitude-modulated light is used so that travel time may be measured in terms of a phase shift observed in the returned modulation envelope. The modulation and demodulation process which is used requires multiple raw frames, which must be captured sequentially, in order to construct a single range image, so accurate recovery of range requires the scene to remain static for the entire period of acquisition. Motion during acquisition causes artifacts to appear.

Optical flow refers to the pattern of motion observed from a particular point of view [2] and is usually estimated using data collected from 2D cameras. Early methods for estimating optical flow utilize spatiotemporal derivatives in order to compute velocity [3]. Assuming that scene illumination is constant, it can be asserted that any local change in image intensity must be due to motion. Common techniques for estimating optical flow use this brightness constancy constraint in combination with other constraints such as an image smoothness constraint in order to enable velocity to be estimated using iterative optimization [4]. The discrete wavelet transform can be used to efficiently form an accurate approximation of any real function, so wavelet analysis can be used to reformulate the estimation of optical flow as the solution to a linear system of equations [5].

Another family of techniques uses the response of spatiotemporal filters in order to estimate optical flow [3, 6]. The dual-tree complex wavelet transform, like all wavelet transforms, essentially comprises a bank of filters. Recognizing that, similar to a Fourier transform, the phase of dual-tree complex wavelet transform coefficients varies approximately linearly with displacement in the spatial domain, an algorithm can be developed which uses these coefficients to estimate optical flow [7]. It has been noted that local phase analysis through the Gabor wavelet transform is robust to local brightness changes [8], but the analysis has so far been limited to classical imaging conditions.

The modulation and demodulation process used in ToF imaging means that the brightness consistency constraint mentioned above is not valid, so adaptations have to be made. For example, dual-tap ToF sensors enable two raw frames with phase shifts π radians apart to be captured simultaneously. In normal ToF camera operation one of these taps is subtracted from the other, greatly mitigating signal from background light. Adding these taps instead provides something akin to greyscale imaging which does meet the requirement for brightness constancy and therefore enables the application of standard optical flow techniques [1, 9].

In this work we find that applying a Fourier transform in time to motion-contaminated time-of-flight range imaging data reveals small waves whose spatial frequency vector encodes velocity information. We therefore propose characterization of these waves using the continuous wavelet transform (CWT) as a means of measuring motion in time-of-flight range imaging.

II. THEORY

A. Amplitude-Modulated Continuous Wave Time-of-Flight Range Imaging

Time-of-flight range imaging systems consist of a light source and a camera, typically placed close together. Modulating the amplitude of the light source by a periodic signal, the time it takes light to travel from the source to the scene and back to the camera induces a phase shift which is given by

$$\phi = \frac{4\pi f d}{c}, \tag{1}$$

where d is the distance between the scene and the camera, f is the frequency of the modulating signal, which is usually in the tens to low

Corresponding author: L. Streeter (e-mail: streeter@waikato.ac.nz).
Digital Object Identifier 10.1109/LSENS.2017.0000000

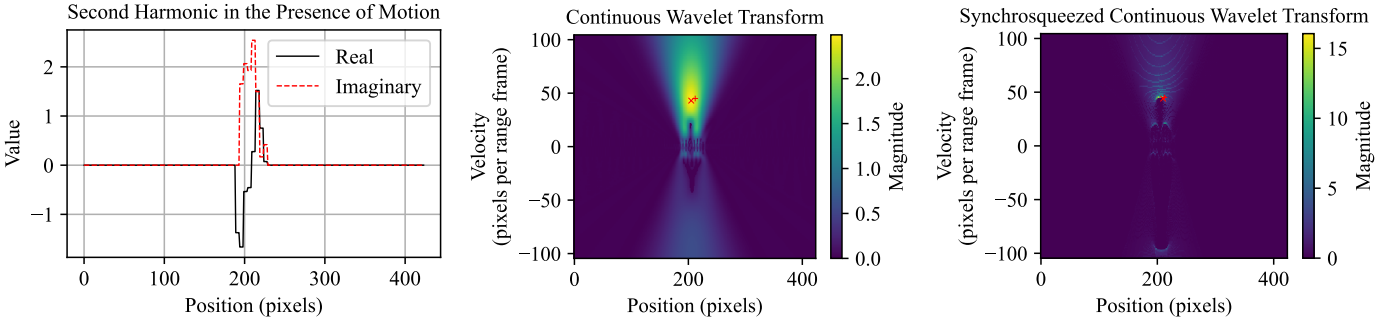

Fig. 1. (left) Second harmonic (second Fourier bin) resulting from the Fourier transform of simulated ToF data in which there is a single edge located at pixel 212 which has a velocity of 45 pixels per range frame. The second harmonic is zero in the absence of motion. (centre) CWT used to find the location and velocity of a moving edge in simulated ToF data. The position of the peak value, denoted by a red "×," indicates that the measured position of the edge is pixel 205 and the measured velocity is 43 pixels per range frame. (right) Synchrosqueezing the CWT. The position of the peak value indicates that the measured position of the moving edge is pixel 210 and the velocity is 44 pixels per range frame.

hundreds of megahertz, and c is the speed of light [1]. Demodulating the returned light in homodyne at every pixel on the sensor yields the so-called correlation waveform:

$$I_n = \alpha \cos(\phi + \theta_n) + \beta, \tag{2}$$

where α is dependent on the amplitude of the returned light, β is a DC offset which is dependent on ambient light, and θ_n is a programmable phase shift between the modulating and demodulating signals [1]. Because (2) contains three unknowns, at least three I_n must be taken with different values of θ_n in order to independently recover α, β and ϕ and subsequently calculate the distance to each point in the scene using (1). Typically, θ_n are spaced evenly by $\Delta\theta$ between 0 and 2π. Taking the discrete Fourier transform of the I_n values over n then enables the recovery of ϕ by examination of the first Fourier bin [10]. Assuming α, β and ϕ each remain the same for all I_n, energy in the discrete Fourier transform is confined to the zeroth bin (caused by the DC term β) and the first Fourier bin and its conjugate (caused by the sinusoidal term).

B. A Model of Transverse Motion in Time-of-Flight Range Imaging

The values of I_n are typically captured sequentially. Motion during acquisition may cause α, β and ϕ to change, meaning energy in the Fourier transform spreads to harmonics that would otherwise be zero (up to noise) [11]. Assuming motion is linear, we can model motion by introducing a motion vector, (u, v), via

$$I(x, y)_n = I(x - nu, y - nv)_{n,u=0,v=0}, \tag{3}$$

where (x, y) denotes position and $I(x, y)_{n,u=0,v=0}$ are the raw values which would have been captured in the absence of transverse motion. Taking the discrete Fourier transform from n to ω and applying the Fourier shift theorem then yields

$$\hat{I}(x, y)_\omega = \hat{I}(x, y)_{\omega,u=0,v=0} e^{-jx\frac{\omega-\Delta\theta}{u}} e^{-jy\frac{\omega-\Delta\theta}{v}}, \tag{4}$$

where $j = \sqrt{-1}$ is the imaginary unit [12]. A shift through space over time is encoded as a shift in phase over frequency. Critically, this phase shift causes small waves to appear which are oriented in the direction of motion and whose spatial frequency is inversely proportional to velocity. Moreover, these waves appear in Fourier bins which have zero energy in the absence of motion. We therefore propose measurement of these waves as a means of recovering optical flow in time-of-flight range imaging.

III. METHODS

A. Optical Flow Estimation in Time-of-Flight Range Imaging Using the Continuous Wavelet Transform

Since motion gives rise to spatial waves, we aim to measure the spatial frequency of these waves using the CWT [13] over each row and column separately of the Fourier transformed images. In the x direction the CWT is:

$$(W_\psi \hat{I}(x, y)_\omega)(a, b) = \int_{-\infty}^{\infty} \hat{I}(x, y)_\omega \frac{1}{\sqrt{a}} \psi^* \left(\frac{x - b}{a}\right) dx, \tag{5}$$

where the superscript $\cdot^*$ denotes complex conjugation. Throughout, we use the complex Gaussian wavelet [14],

$$\psi(x) = \sqrt{2\pi} s^2 e^{\frac{-s^2x^2}{2} + j\mu x}, \tag{6}$$

and select $\mu = 2.35$ and $s = 0.92$ (by inspection). Both scale, a, in (5) and velocity, u, in (4) are inversely proportional to spatial frequency, so a is proportional to velocity. Shift, b, in (5) enables us to localize motion artifacts in space. Finding the maximum of the magnitude of the CWT identifies the position and velocity of a single moving object in a row (or column). (Whereas by finding the maximum at over a at each position, b, of the CWT we can obtain an estimate of velocity at each pixel.)

We operate on the second harmonic of ToF data as it is the lowest harmonic that contains energy in the presence of motion but no energy for static scenes. The sign of a provides information about direction, so the second harmonic must be complex in order for us to distinguish between positive and negative velocities. We must therefore use at least five raw frames per range frame; four raw frames per range frame are sufficient to return a second harmonic, however it is purely real, resulting in a CWT which is symmetric in a, preventing positive and negative velocities from being distinguished from each other.

In Fig. 1 we show a one-dimensional simulation of the proposed optical flow method. Note that while ToF imaging ordinarily returns depth estimates across a two-dimensional field, our simulation is equivalent to one row or column on a ToF sensor. The "scene" contains a single moving edge, and the second harmonic of the Fourier transform in time exhibits a single period of a wave in space, as predicted by Eqn. 4. Performing a spatial CWT, the peak value's location in the b (shift) axis indicates the predicted position of the

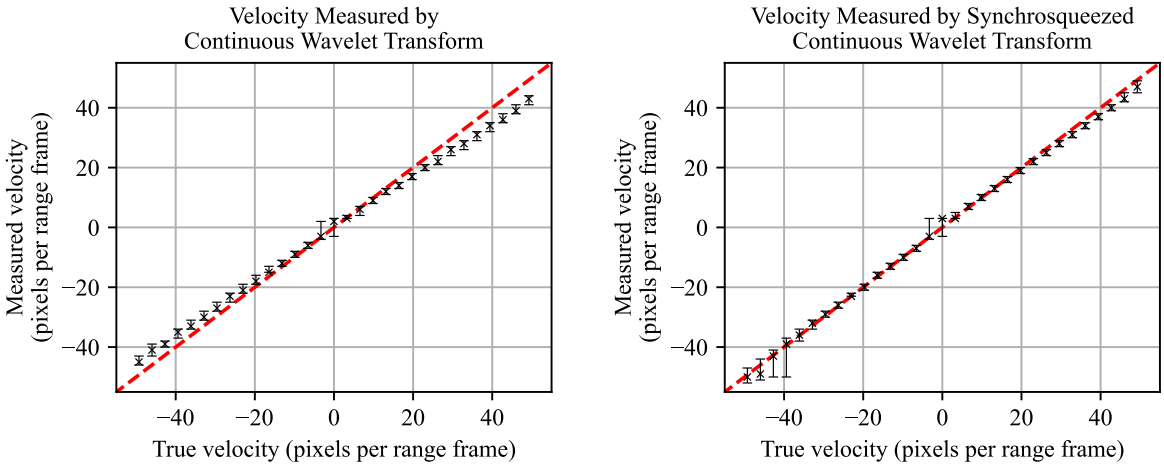

Fig. 2. Experimental velocity measurement with nine raw frames per range frame at a distance of three metres using (left) the CWT only and (right) synchrosqueezing. At each true velocity, the median estimated velocity is plotted as an "x" and the errorbars indicate the interquarile range of the estimated velocities. The red line is the reference true velocity line.

edge and its location in the a (scale) axis indicates the predicted velocity of the edge.

B. Synchrosqueezing

The CWT produces a strong signal near the true position and velocity of moving edges. However, the peaks are broad. Synchrosqueezing usefully sharpens the peaks, concentrating energy in the scale axis [15] (which is proportional to velocity). Synchrosqueezing dramatically sharpens the peaks in the velocity axis of the wavelet transform (see Fig. 1).

C. Evaluation and Testing

We test our methods using a selection of different numbers of raw frames per range frame (odd numbers from 5 to 13) with the camera at positioned separately at one, two and three metres from the scene. We capture data using a Chronoptics Kea ToF camera (Chronoptics Ltd., Hamilton, New Zealand.) The modulation frequency is 50 MHz, so the ambiguity distance is 3 m (e.g. the phase shift observed at 1 m is the same as at 4 m), meaning that each distance tested corresponds to a unique phase shift. In each case, the scene consists of a single, large, diffusely reflective rectangular target against a dark background. The target is moved transverse to the camera to a number of equally spaced positions on a horizontally oriented translation stage. A complete set of raw frames is collected at each position and motion is synthesized by selecting raw frames such that the rectangle moves a known increment between raw frames. Velocities of up to 50 pixels per range frame are used. Each row of the frame is processed independently and only one velocity is detected per row. The location of the motion is taken as the position of the maximum of the magnitude of the CWT, as in Fig. 1. Note that while this approach does take account of motion between raw frames (inter-frame blur) it does not take account of motion which occurs during acquisition of each individual raw frame (intra-frame blur).

IV. RESULTS AND DISCUSSION

We present the mean absolute error results from our tests in Table 1, and a graph of the estimated velocity at three metres distance and nine raw frames in Fig. 2. Overall, we find that error values are good, typically no more than five pixels per range frame (10% of the maximum true velocity tested). Additionally, except for the trials where the five raw frames are used and where the distance between the camera and the scene is two metres, synchrosqueezing reduces mean absolute error. This is particularly evident in the trials in which the distance between the camera and the scene is three metres, where mean absolute error achieved by using synchrosqueezing is as little as one third that of the continuous wavelet transform on the same data. As less light is returned at longer distances, we speculate that the improved performance of synchrosqueezing is because it concentrates of energy in the scale (velocity) axis, improving the effective signal-to-noise ratio of the CWT.

While the work above includes spatial localization of the velocity of moving objects, we are only finding one moving object. As stated above, finding the peak in the a axis for each position in the b axis of the CWT, we are able to estimate velocity at every pixel in a row. Performing this process on every row and column in a ToF image then yields an estimate of horizontal and vertical velocity for each pixel, enabling us to demonstrate simple optical flow estimation in Fig. 3. A threshold on the magnitude of the second harmonic forms a per-pixel motion mask, used to set the velocity of pixels without motion to zero. ToF data are captured of a cow moving from right to left in the scene. The horizontal estimates are mostly correct, but small objects with high velocity cause erroneous measurements. As the true motion is mostly in the horizontal direction, the vertical estimates suffer from the so-called aperture problem. The thin regions of non-zero optical flow in Fig. 3 are reminiscent of the seminal single-scale optical flow attempts [4]. The scale adaptation naturally provided by the wavelet transform has unfortunately not translated into implicit multi-resolution analysis.

TABLE 1. Mean absolute error for transverse velocity measurements taken at three different distances and five different numbers of raw frames per range frame using both the standard continuous wavelet transform (CWT) method, and the method augmented with syncrosqueezing (SS). Velocities and error values are measured in pixels per range frame.

Distance:	1 m		2 m		3 m	
Method:	CWT	SS	CWT	SS	CWT	SS
5 Raw Frames	2.2	2.8	4.5	3.5	24.0	10.8
7 Raw Frames	3.9	2.9	2.3	3.6	10.7	3.5
9 Raw Frames	4.3	3.3	3.3	4.6	4.4	2.5
11 Raw Frames	5.2	3.9	3.3	4.7	3.6	2.2
13 Raw Frames	6.4	5.0	2.2	3.4	3.7	2.1

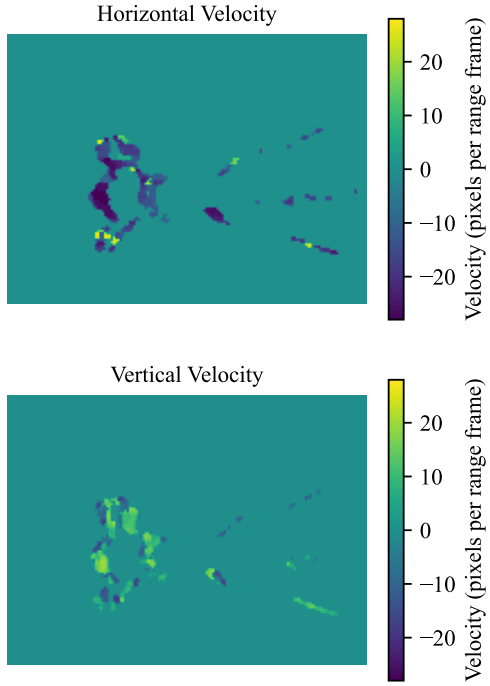

Fig. 3. Velocity measured by the synchrosqueezed wavelet transform in two dimensions. The scene contains a cow walking from right to left. The data are downscaled by a factor of two in the horizontal and vertical directions prior to analysis, but velocities are reported at the original scale.

V. CONCLUSIONS AND FUTURE WORK

Transverse motion occurring during data acquisition in time-of-flight range imaging causes spatial waves to appear in the Fourier transform. The spatial frequency vector of these waves usefully encodes the speed and direction of the motion that occurred. We have successfully shown that the CWT enables us to characterize these spatial waves and thereby estimate the velocity of the motion which caused them.

Full wavelet analysis is not necessarily efficient, particularly in two or more dimensions, so in future work we will investigate faster methods for real-time applications. Additionally, although the process of recovering depth in time-of-flight naturally exposes the artifacts upon which our work relies, the underlying theory does not limit its application to time-of-flight range imaging. The potential of phase based large velocity optical flow in standard camera data will be explored in future work.

ACKNOWLEDGMENT

This research was supported by a New Zealand Ministry of Business, Innovation and Employment Smart Ideas grant (UOWX1705).

REFERENCES

[1] D. Lefloch, R. Nair, F. Lenzen, H. Schäfer, L. Streeter, M. J. Cree, R. Koch, and A. Kolb, *Technical Foundation and Calibration Methods for Time-of-Flight Cameras*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 3–24.

[2] D. Sun, S. Roth, and M. J. Black, "Secrets of optical flow estimation and their principles," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010, pp. 2432–2439.

[3] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, "Performance of optical flow techniques," *International Journal of Computer Vision*, vol. 12, pp. 43–77, 1994.

[4] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, no. 1–3, pp. 185—-203, 1981.

[5] L.-F. Chen, H.-Y. M. Liao, and J.-C. Lin, "Wavelet-based optical flow estimation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 1, pp. 1–12, 2002.

[6] D. J. Heeger, "Optical flow using spatiotemporal filters," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 279–302, 1988.

[7] H. Chen and N. Kingsbury, "Efficient registration of nonrigid 3-D bodies," *IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 262–272, 2012.

[8] D. J. Fleet and A. D. Jepson, "Computation of component image velocity from local phase information," *International Journal of Computer Vision*, vol. 5, no. 1, pp. 77–104, 1990.

[9] M. Lindner and A. Kolb, "Compensation of motion artifacts for time-of-flight cameras," in *Workshop on Dynamic 3D Imaging*. Springer, 2009, pp. 16–27.

[10] L. Streeter, "Time-of-flight range image measurement in the presence of transverse motion using the Kalman filter," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 7, pp. 1573–1578, 2018.

[11] S. Hussmann, A. Hermanski, and T. Edeler, "Real-time motion artifact suppression in ToF camera systems," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 5, pp. 1682–1690, 2011.

[12] L. Streeter and J. J. Wise, "Exploring machine learning to reduce motion error in time-of-flight range imaging," in *Imaging and Applied Optics Congress*. Optical Society of America, 2020, p. DF3A.3.

[13] S. Mallat, *A Wavelet Tour of Signal Processing*. Burlington, Massachusetts, USA: Academic Press, 2009.

[14] G. Thakur, E. Brevdo, N. S. Fučkar, and H. T. Wu, "The synchrosqueezing algorithm for time-varying spectral analysis: robustness properties and new paleoclimate applications," *Signal Processing*, vol. 93, no. 5, pp. 1079–1094, 2013.

[15] I. Daubechies and S. Maes, *A Nonlinear Squeezing of the Continuous Wavelet Transform Based on Auditory Nerve Models*. Boca Raton, Florida, USA: CRC Press, 1996, pp. 527–546.