Adapting the Bidirectional Kalman Filter for use in Multi-Frequency Time-of-Flight Range Imaging

A thesis

submitted in fulfilment

of the requirements for the Degree

of

Master of Engineering

at

The University of Waikato

by

Ali Nouruldaim Alqassab

2020

Abstract

Time-of-flight range imaging cameras obtain range by producing amplitude modulated light and measuring the time taken for light to travel to the scene and back to the camera. Time-of-flight cameras require at least three raw measurements to calculate range. Raw frames are captured sequentially, and as such, motion in scenes during capture leads to inconsistent raw frame measurements and erroneous range calculations. Motion error constrains Time-of-flight cameras to non-dynamic scenes and limits their potential applications. The Time-of-flight bidirectional Kalman filter method is a state-of-the-art method known to reduce error due to transverse motion in cameras operating with a single modulation frequency. The method works by treating raw frames as a noisy time series and running the Kalman filter on it to produce a range estimation at every raw frame. The Kalman filter is then reapplied to the data in reverse to produce another set of range estimations, and a composite range is selected from the two set of range estimations. A number of commercial time-of-flight cameras, such as the Microsoft Kinect V2, use multiple modulation frequencies. In this thesis, we adapt the bidirectional Kalman filter method to multi-frequency operated cameras by having the prediction component of the Kalman filter take into account the change in amplitude and phase shift caused by the change in frequency. The amplitude component of the prediction is performed linearly by multiplication, while the phase shift component of the prediction is performed using the ratio of the modulation frequencies. The phase shift prediction across modulation frequencies requires the phase to be unwrapped. The unwrapping is performed between modulation frequencies by selecting the number of phase wraps that best predicts the two following raw frames. Finally, to ensure correct composite phase selection, an alternative method for selecting the composite phase is proposed for the adapted bidirectional Kalman filter. We perform quantitative and qualitative experiments to test the proposed method. In the quantitative experiment, the proposed method produces less error than the classical Discrete Fourier Transform approach in 70% of tested instances. The qualitative experiment shows that the proposed method significantly reduces motion blur.

Acknowledgements

Firstly, I thank God for his guidance and for giving me the strength and determination that I needed to pursue this study and reach this far.

I would like to express my gratitude and thanks to my supervisor, Dr. Lee Streeter, for his expertise, patience, and enthusiasm throughout this research project. His door was always open to discuss ideas and problems, and his support was invaluable.

I also want to thank my co-supervisors, Dr. Michael Cree and Dr. Shen Hin Lim, for their feedback.

For the many useful discussions we had, I'd like to express my gratitude to Vance, Gehan, Carl, Hedi, and Dale. Our discussions were always interesting and insightful.

To my friends, the Faisal and Al Ghrabi families, I offer my thanks for their encouragement along the way. Finally, I'd like to thank my family for their unwavering support. I would like to give special thanks to my parents for the sacrifices they have made to allow me to pursue my academic endeavours.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Full-field range imaging cameras provide depth measurements of the imaged scene. Several range imaging techniques exist. Stereoscopic imaging, for example, uses a triangulation technique similar to that of the human visual system. Two sensors, separated by a known distance, are used to image a scene. Range is then calculated using the offset of objects observed between the two sensors. Structured light is another range imaging technique. The scene is illuminated by infrared light comprised of many dots and an algorithm uses the disparity of the imaged dots to calculate depth. Time-of-flight range imaging cameras, unlike other range imaging techniques, are able to provide accurate depth measurements without the need for large amounts of computational power. Time-of-flight range imaging cameras use a light source to illuminate a scene. The light then backscatters into the camera, and the time taken for the light to travel to the scene and back to the camera is used to calculate the distance traveled by the light. Time-of-flight cameras have become a vital instrument in many fields such as robotics, where depth information is needed for accurate navigation, and in gaming, where human skeletal information is recognised and tracked. The potential uses of time-of-flight cameras, however, are limited by errors inherent to them. A significant source of error that affects time-of-flight cameras is motion. Motion during capture is especially significant in time-of-flight because, unlike other imaging techniques, time-of-flight requires multi-

ple sequential measurements to obtain a single range image. As such, motion during capture leads to inconsistent measurements, and thus erroneous range calculations. However, in recent years, several methods have been developed to reduce motion error and provide more accurate range measurements. Other forms of error exist in time-of-flight imaging, such as multi-path interference and phase wrapping. The effect of these errors on range imaging is mitigated by techniques involving the use of multiple modulation frequencies.

1.1 Objective

Although a vast number of commercial time-of-flight range imaging cameras use multi-frequency operation to obtain range, none of the motion error reduction techniques developed are designed to work with or have been adapted to work with multi-modulation frequency operated cameras. The objective of this thesis is to adapt a recently developed state-of-the-art motion reduction technique called the time-of-flight bidirectional Kalman filter. The adaptation to multi-frequency operation requires an understanding of the changes that occur to the data across frequencies and how they affect the time-of-flight bidirectional Kalman filter. The question which this thesis attempts to answer is: how do we adapt the time-of-flight bidirectional Kalman filter method to a multi-modulation frequency time-of-flight camera?

1.2 Thesis structure

In **chapter two**, we first revise the time-of-flight theory. We then introduce the sources of error in time-of-flight and the methods used to reduce them. The types of motion error are introduced, and a literature review of motion error reduction techniques is presented. Finally, a thorough review of the time-of-flight bidirectional Kalman filter method is provided.

In **chapter three**, we adapt the bidirectional Kalman filter to the time-of-flight multi-frequency operation in static scenes. The adaptation is performed

by taking into account the changes that occur to the data between modulation frequencies and performing the changes necessary for an accurate bidirectional Kalman filter phase estimation. The change that occurs to phase shift and amplitude caused by the change in modulation frequency is discussed, and the adapted method is proposed. Experiments are performed to test the proposed adapted method for the amplitude and phase shift across modulation frequencies as well as the overall performance of the proposed adapted bidirectional Kalman filter on static scenes. The results of these experiments are then presented and discussed.

In **chapter four**, we test the adapted bidirectional Kalman filter proposed in the previous chapter on dynamic scenes and changes are made to the model to enable transverse motion error reduction in dynamic scenes. We propose an alternative method for selecting the composite phase shift and perform a quantitative experiment to test whether the proposed composite phase selection method improves accuracy. Finally, we conduct qualitative and quantitative experiments to test the overall effectivness of the proposed adapted bidirectional Kalman filter method on dynamic scenes. The results are compared to the unadapted method and the traditionally used discrete Fourier transform method.

In **chapter five**, we conclude the thesis and present an overview of the proposed method and the results. Limitations of the proposed method is discussed and possible future work is presented.

1.3   Publication arising from this thesis

The work presented in this thesis has resulted in the following publication:

A. N. Alqassab, L. Streeter, M. J. Cree, C. A. Lickfold, V. Farrow, and S. H. Lim. Adaptation of Bidirectional Kalman Filter to Multi-Frequency Time-of-Flight Range Imaging. In: Image and Vision Computing New Zealand (IVCNZ 19). IEEE (2019).

Chapter 2

Background

2.1 Time-of-Flight Principle

Time-of-flight cameras contain a light source that illuminates the scene with infrared light. The time taken for light to travel from the source to the scene and back again is used to measure range. There are two methods for measuring the time-of-flight: the pulsed direct time-of-flight method, and the amplitude modulated continuous wave (AMCW) method. Pulsed time-of-flight cameras calculate range by measuring the time taken for the pulse to go to the scene and return back to the camera. In the AMCW method, which is the method most commonly used by commercial time-of-flight cameras (Grzegorzek *et al.* (2013)), the camera illuminates the scene with modulated light. The travel time causes a phase shift in the reflected signal which is used to calculate range. The light produced by the AMCW camera is modeled by

$$s\left(t\right) = \cos\left(2\pi f_m t\right), \tag{2.1}$$

where f_m is the light modulation frequency. The infrared light backscatters from objects in the scene to the camera and is captured by a sensor with a high-speed modulated shutter. At the receiving end, as shown in Fig. 2.1, the reflected signal undergoes an additional phase shift ϕ.
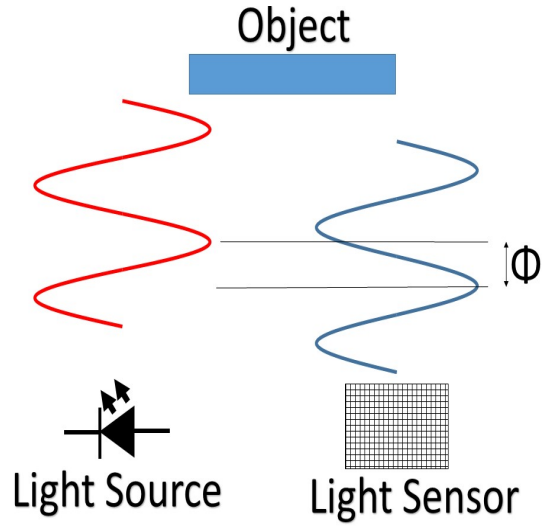
Figure 2.1: Time-of-flight phase shift.

The phase shift induced depends on the distance travelled, specifically,

$$\phi = \frac{4\pi f_m d}{c},\tag{2.2}$$

and the reflected signal is given by,

$$g\left(t\right) = A\cos\left(2\pi f_m t + \phi\right) + B,\tag{2.3}$$

where ϕ is the phase shift, B is the DC offset, A is the amplitude, and c is the speed of light.
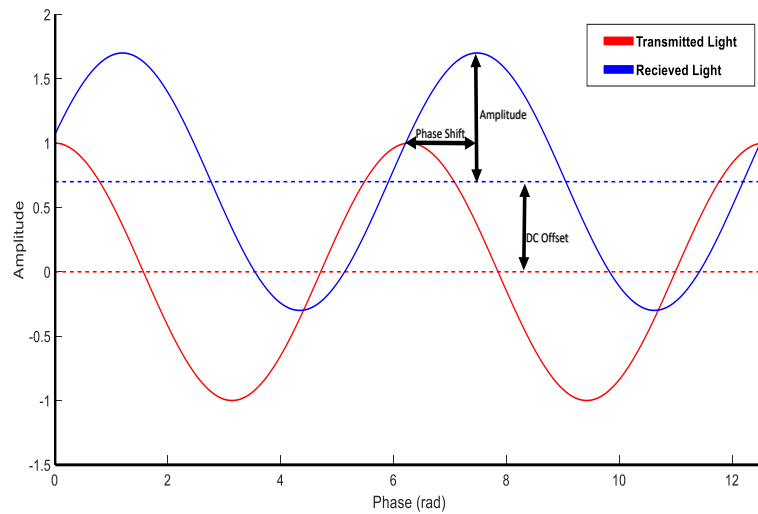
Figure 2.2: Time-of-flight amplitude, phase Shift and DC offset.

The sensor is shuttered in homodyne with the light source, demodulating the returned light. The demodulated signal is integrated with a programmed phase offset, θ, between the light source and sensor. The correlation of the reflected signal and the shuttered signal occurs on chip with the two signals mixed, and is modelled by,

$$I(\theta) = \lim_{T \to \infty} \int_{\frac{-T}{2}}^{\frac{T}{2}} s(t + \theta) g(t) \, dt. \tag{2.4}$$

Trigonometric expansion and evaluation of the integral above is simplified by Lange and Seitz (2001) to the equation

$$I(\theta) = \frac{A}{2} \cos(\theta + \phi) + B. \tag{2.5}$$

At least three measurements at different phase offsets are required to solve for the phase shift, DC offset, and amplitude. Fixing $\theta = \theta_n$, a measurement I_n at a particular phase offset is referred to as a raw frame,

$$I_n = \frac{A}{2} \cos(\theta_n + \phi) + B. \tag{2.6}$$

The phase offsets are equally spaced values from 0 to 2π. Taking the discrete Fourier transform (DFT) of the raw frames, the phase shift ϕ is found from the the first Fourier bin (Streeter (2018)). The DFT is the traditional method used to calculate the phase shift from the raw frames.

2.2   Sources of Time-of-Flight Error

AMCW time-of-flight cameras are affected by a variety of errors which lead to inaccurate range measurements. Some of these errors are caused by external factors from the environment, while others are intrinsic to the cameras. Several papers exist which review these errors and the methods used to reduce them. Foix *et al.* (2011) review some of these time-of-flight errors as well as early calibration methods. He and Chen (2019) review more recent methods that reduce time-of-flight errors. The paper, however, does not discuss more recent methods in motion error reduction, such as the stochastic calculus method (Streeter (2017b)), and the bidirectional Kalman filter method

(Streeter (2018)), which will be discussed in further detail below. Streeter and Kuang (2019) also review the time-of-flight errors, as well as older and more recent methods that reduce them. Below, we review some of the errors that affect time-of-flight cameras and some of the methods proposed to reduce them.

2.2.1 Harmonic error

Harmonic error, sometimes referred to as the wiggling error, is a systematic error that occurs because the equations used to determine the phase shift are based on idealised sinusoidal waves. In reality, the light source and the sensor shutter signals are produced by digital electronics using square waves, and the actual signals result in a triangular correlation signal containing odd order harmonics. Error due to aliasing can occur when sampling at a frequency less than twice that of the highest harmonic in the correlation signal. The harmonic error is distance dependant, which results in a depth measurement which oscillates around the true depth. One way to solve this is by using
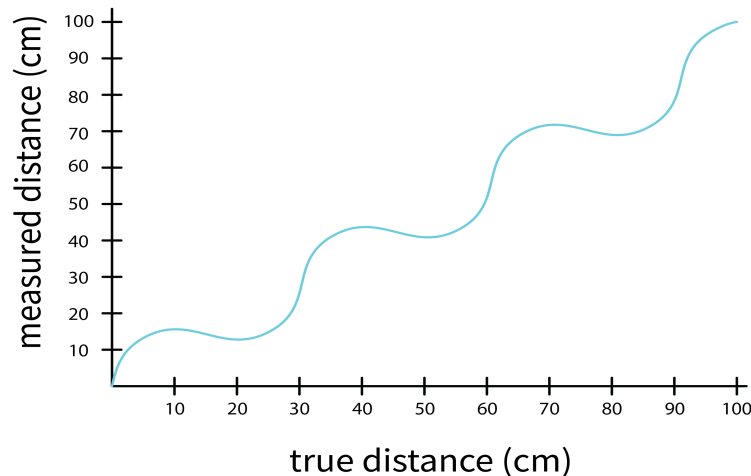
Figure 2.3: Time-of-flight harmonic error.

equations that take into account the higher harmonics. This solution, however, demands a greater number of sampling points and more calculations from the

time-of-flight cameras (Dorrington *et al.* (2008)). A simpler approach is to use a calibration method to estimate and subtract the harmonic error (Grzegorzek *et al.* (2013)). A different approach to reducing the harmonic error involves changing the process of acquisition to reduce the error induced by the third harmonic, which contributes the most to the harmonic error (Streeter and Dorrington (2015)),(Payne *et al.* (2008)).

2.2.2 Multi-Path Interference Error

In an ideal scenario, each pixel in the time-of-flight sensor receives an amplitude modulated sinusoidal light wave reflected from a single point in the scene. When the light reaches the pixel from multiple locations, it results in what is known as multiple return error. This error can be categorised into two types. The first type is known as mixed pixel error, which occurs when imaging the edge of an object, and light arriving to the pixel is made up of both the object and the background, as shown in Fig. 2.4.
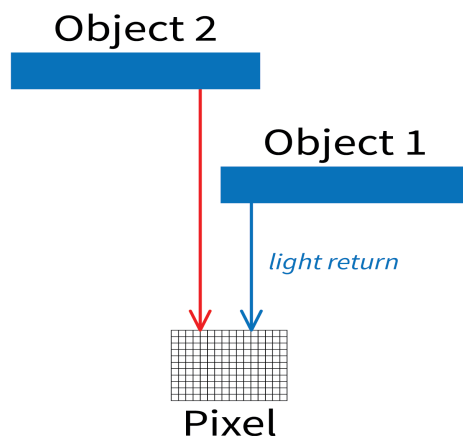
Figure 2.4: Time-of-flight mixed pixel error.

The second type of error is known as the multi-path interference error, and it occurs when light travels multiple paths to reach a pixel. One form of multi-path interference is scene based, as shown in Fig. 2.5, where, by reflecting off objects within the scene, the light travels along multiple paths before it is received by the pixel.
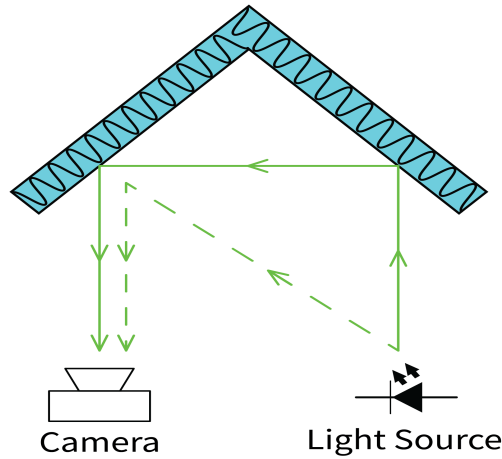
Figure 2.5: Time-of-flight multi-path interference error.

A second, camera intrinsic, source of multi-path interference is internal scattering. The scattering occurs when a portion of the incident light reflects back and forth between the sensor and lens. These internal reflections cause the pixels to receive light not only from the object, but also from other pixels that receive light from different areas in the scene, as shown in Fig. 2.6. The signal at the pixel, therefore, becomes a sum of the phase vectors of those various signals. This is why the effect of the error on the measurement is more prominent when the object in the foreground is more reflective than the background, and if the object is much closer to the camera than the background (Karel *et al.* (2012)).
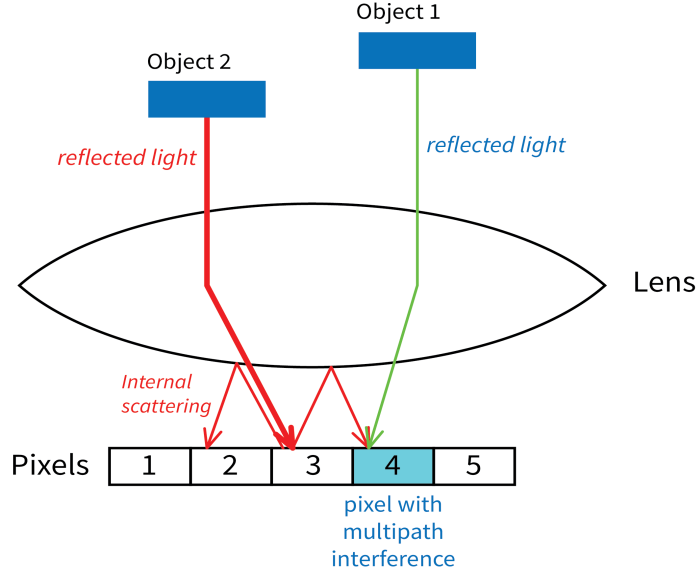
Figure 2.6: Time-of-flight internal scattering.

In an idealised scenario, wherein light travels along a single path from the light source to the scene and back to the camera, it can be expressed as a complex phasor.

$$\eta = Ae^{(j\phi)} \tag{2.7}$$

where A is the amplitude and ϕ is the phase shift. When the pixel receives multiple returns of light originating from different parts of the scene, the return that is meant to be received by the pixel, which usually has the highest amplitude, is called the primary return. Any other return that arrives to the pixel is a secondary return.

Since the pixel receives multiple signals, the pixel measurement is expressed as a summation of N returns, as shown in Eqn. 2.8,

$$\xi = \sum_{n=1}^{N} \eta_n = \sum_{n=1}^{N} A_n e^{(j\phi_n)} \tag{2.8}$$

where N is the total number of returns, η_n is the phasor of return n. There are multiple ways of reducing the multi-path error. Fuchs (2010) attempts
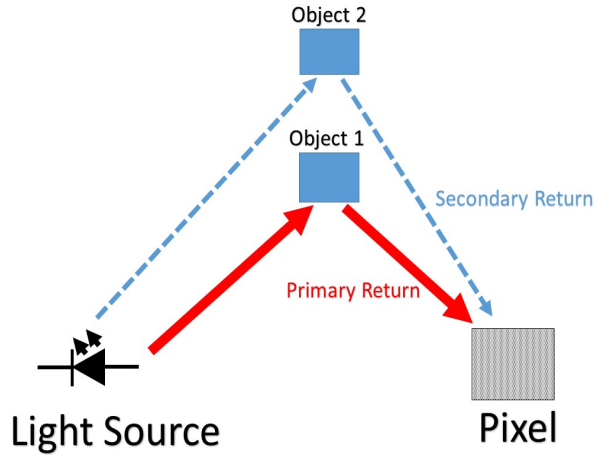
Figure 2.7: Multiple returns in multi-path interference.

to reduce the impact of interference by creating a model which assumes that light in a scene perfectly diffuses (lambertian reflectance), the model is then used to estimate and correct the error from multi-path interference. Godbaz *et al.* (2012) reduce the multi-path error by taking four measurements with a different modulation frequency for each measurement. The results are then used to solve for two returns.

2.2.3 Phase ambiguity due to Wrapping

We measure range based on the phase shift of a sinusoidal wave. The maximum phase shift that can be measured is 2π. The distance that corresponds to this maximum phase shift is referred to as the maximum ambiguity distance. The maximum ambiguity distance, L, is dependent on the modulation frequency, and can be calculated using the following equation ((Grzegorzek *et al.*, 2013, pp. 5-6)):

$$L = \frac{c}{2f_m}. \tag{2.9}$$

An object beyond the maximum ambiguity distance causes the phase shift to wrap around, making it appear closer than it actually is. Due to this phenomenon, a measured phase shift could correspond to several distances depending on the number of times the phase shift has wrapped, as shown in the Fig. 2.8.
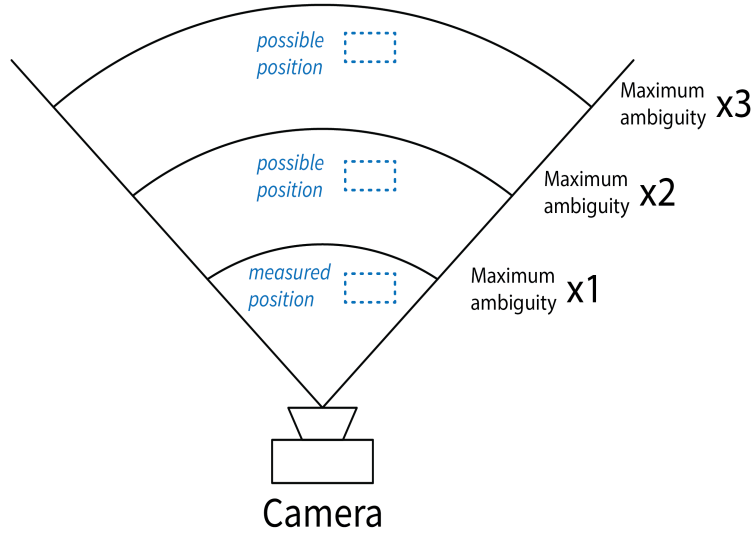
Figure 2.8: Phase ambiguity due to wrapping.

Since phase wrapping occurs beyond the maximum ambiguity range, the unambiguous range of the camera is maximized by reducing the modulation frequency of the time-of-flight camera. However, a lower modulation frequency is more vulnerable to random noise and results in measurements with lower accuracy (Lange (2000)).

A different solution to this problem is to capture range images using multiple modulation frequencies, as seen in Fig. 2.9, each image provides a set of possible range values (Bamji *et al.* (2015)). The value that is present in both modulation frequencies is chosen to be the true range.

This method does not get rid of all ambiguity, but extends its range according to Eqn. 2.10.

$$L = \frac{c}{2|f_{m1} - f_{m2}|}.$$
(2.10)

where L is the maximum ambiguity distance, c is the speed of light, f_{m1} and f_{m2} are the modulation frequencies.
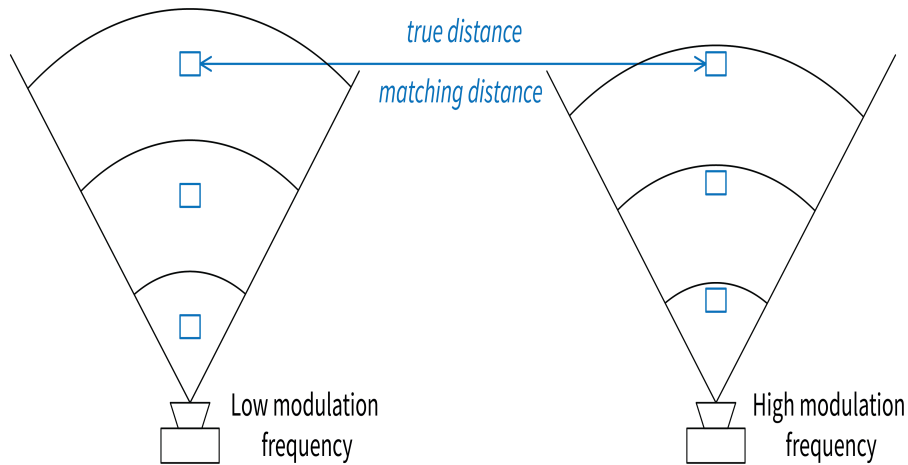
Figure 2.9: Phase unwrapping using multiple modulation frequencies.

2.2.4 Motion Error

The motion of objects during the range measurements of time-of-flight cameras causes inaccurate range calculations. This error is different to that of motion blur in regular two dimensional imaging cameras, where, between background and foreground, a smoothing of colour occurs. Instead, time-of-flight motion error causes an underestimation or an overestimation of the range measurement. Motion error occurs in time-of-flight because multiple raw frames are required to obtain a single range image, and since the process of obtaining multiple frames does not occur simultaneously, any motion that occurs between raw frames results in raw frame measurements corresponding to different amplitudes, phase shifts and DC offsets. Since each raw frame has its own integration time, motion that occurs during the integration time causes what is known as intra-frame motion error, while error due to motion between raw frames is known as inter-frame motion error. Time-of-flight error due to motion is categorised into two types: error due to transverse motion, and error due to radial motion. Transverse motion occurs when objects move in a direc-

tion perpendicular to the field of view of the camera. In this type of motion, an object that passes over the field of view of a pixel causes it to measure raw frames of the background and then of the foreground object as it passes over the pixel's view. Transverse motion also results in a pixel viewing the foreground object and then the background to incur error. As such, the error mostly occurs at the edges of objects.
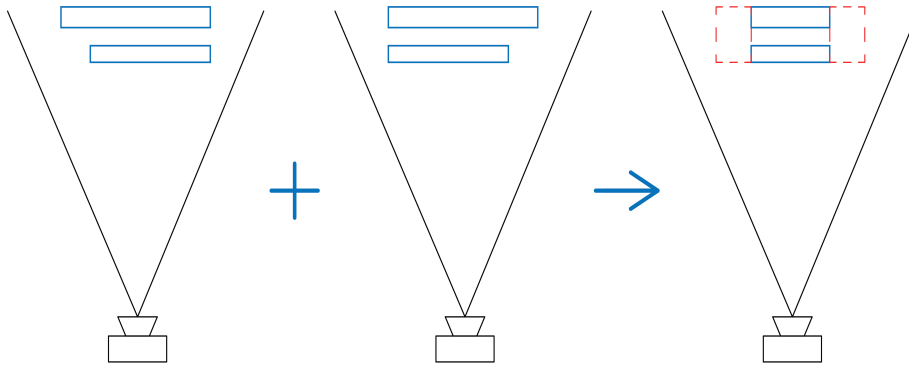
Figure 2.10: Transverse motion error. The figure shows a background object and a foreground object moving in transverse fashion illustrating that when transverse motion occurs, regions around the edges of the moving object are more prone to transverse motion error.

In radial motion, the object moves directly towards or away from the pixel. This leads to range images calculated from raw frames that reflect different phases shifts. As shown in Fig. 2.11, when an object moves in radial fashion, radial motion effects all the pixels viewing the object, not just the boundaries.
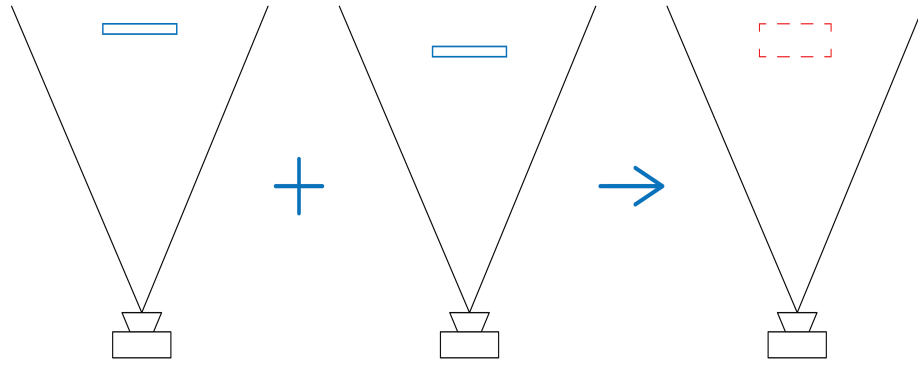
Figure 2.11: Radial motion error. The figure shows an object moving in radially, illustrating that when radial motion occurs, the inconsistencies in the raw frame images effects range measurement of the entire moving object.

Reducing the integration time (time required for the sensor to photo-charge), or reducing the number of raw frames required to compute a range image is known to reduce the motion artefacts (Streeter (2017a)). A number of other methods have been developed to detect and reduce motion error.

Some early methods used to reduce motion error in time-of-flight imaging include combining a colour camera and time-of-flight camera. Since 2D images only require one frame, they are less prone to motion error than time-of-flight imaging (Lottner *et al.* (2007)). Time-of-flight pixels with motion error are identified by performing an edge detection algorithm on a 2D image and finding the corresponding pixels in the time-of-flight image (Lottner *et al.* (2007)).

Lindner and Kolb (2009) model time-of-flight motion blur using optical flow analysis of the raw images for transverse motion and radial motion. The optical flow method enables the range image computation in a pixel to use raw frames from surrounding pixels to decrease motion error artefacts. In order to achieve this, the brightness of corresponding surfaces are assumed to be the same as the preceding images. This method also assumes pixel homogeneity by performing an intensity-value standardisation calibration. The optical flow method, however, is computationally intensive. The performance of the optical

flow method was later improved by Lefloch *et al.* (2013), where, instead of computing three optical flows, the method only estimates two and derives a third optical flow from them.

Streeter and Dorrington (2014) also built upon the optical flow method by combining it with coded exposure. The method allowed for the correction motion within each raw frame as well as between the frames.

Jimenez *et al.* (2014) developed a method which utilises the fact that a range image should be calculated from consistent raw frames. The method first performs a calibration which enables the detection of inconsistencies in the raw frames. Once an inconsistency event is detected, it is characterised as either a falling or rising depth transition. Finally, neighbouring pixels are used to replace the inconsistent data affected by the transverse motion. Instead of reducing the effect of radial motion by reducing the number of raw frames or the integration time, Streeter (2017a) developed methods which treat the impact of motion on the data as a desirable trait that enables the measurement of linear radial velocity. Two different methods were developed to estimate velocity: the Pseudo-Quadrature Signal Analysis method, which enables low velocity estimations, and a correlation analysis method, which performs better at higher velocities. Both methods enable the correction of the error due to radial motion.

Streeter (2017b) also used stochastic calculus to analyse time-of-flight data to develop linear radial estimation methods. One method uses the maximum likelihood estimation of a distribution obtained from stochastic calculus. Another method, named ratio of amplitude, provides a closed-form and more efficient solution.

Another method developed by Streeter (2018) is the bidirectional Kalman filter method, where the raw frames from the camera are treated as noisy time series. The Kalman filter is applied such that an amplitude, phase shift, and DC offset are predicted for every raw frame. The Kalman filter is reapplied to the raw frames in reverse order to provide a second set of predictions. The

final range is chosen from the set that provides the least prediction error. The bidirectional Kalman filter method was proven to reduce transverse motion error in 80% of pixels without requiring additional computational hardware.

None of the methods mentioned above were adapted or designed to work on a multi-modulation frequency time-of-flight operation. Many off-the-shelf time-of-flight cameras, such as the Microsoft Kinect V2, support a multi-frequency operation and are still affected by motion error. This makes it important to adapt a state-of-the art method to a multi-frequency time-of-flight camera. To adapt the time-of-flight bidirectional Kalman filter method to a multi-modulation frequency time-of-flight camera, a detailed study of the bidirectional Kalman filter method is required.

2.2.5 Bidirectional Kalman Filter method

The Kalman filter is a well known statistical analysis filter developed by Rudolf Kalman in 1960. It is widely used in control applications to provide accurate estimations given a noisy time series set of measurements. The Kalman filter is a recursive process, where at each iteration, *a posteriori* is estimated based on the measurement and *a priori* estimation. The Kalman gain, which is calculated using the covariance matrix of the *a priori* and the variance of the measurements, is used to determine whether the model places more weight on the measurements or the *a priori* when estimating the *a posteriori*.

To apply the Kalman filter to time-of-flight data, a time-of-flight linear system is required. The running time-of-flight method is a known time-of-flight linear system that is used as an alternative to the discrete Fourier transform method. Running time-of-flight enables range to be estimated using any three consecutive raw frames. Running time-of-flight is derived by performing a trigonometric expansion on Eqn. 2.6 and the raw frame rewritten as (Streeter *et al.* (2013)),

$$I_n = A \left(\cos \left(\theta_n \right) \cos \left(\phi \right) - \sin \left(\theta_n \right) \sin \left(\phi \right) \right) + B \tag{2.11}$$

$$= \begin{bmatrix} \cos \left(\theta_n \right) & -\sin \left(\theta_n \right) & 1 \end{bmatrix} \begin{bmatrix} A \cos \left(\phi \right) \\ A \sin \left(\phi \right) \\ B \end{bmatrix} = H_n X_n \tag{2.12}$$

where $H_n = \begin{bmatrix} \cos \left(\theta_n \right) & -\sin \left(\theta_n \right) & 1 \end{bmatrix}$, and $X_n = \begin{bmatrix} A \cos \left(\phi \right) & A \sin \left(\phi \right) & B \end{bmatrix}^T$. In this form, X_n is calculated from a series of consecutive raw frames, while phase, amplitude, and DC offset is determined from X_n.

The time-of-flight Kalman filter is a recursive time series process that attempts to estimate a more accurate value of X_n (Kalman (1960)). At each iteration, the filter first predicts the state matrix X_n and the process covariance matrix P_n. Assuming a static scene, the amplitude, phase shift, and DC offset are expected to be the same. As such, the *a priori* prediction of X_n is simply the previously predicted X_n with process noise ϵ_n

$$X_n^- = X_{n-1} + \epsilon_{n-1}, \tag{2.13}$$

From this we can also estimate the raw frame measurement I_n at any given phase step, including sensor noise γ_n, can be found.

$$I_n = H_n X_n + \gamma_n, \tag{2.14}$$

The predicted *a priori* covariance P_n^- is calculated as

$$P_n^- = P_{n-1} + Q, \tag{2.15}$$

where Q is the process noise covariance matrix.

The next step in the Kalman filter is to calculate the Kalman gain, which is

$$K_n = \frac{P_n^- H_n^T}{H_n P_n^- H_n^T + r_n}, \tag{2.16}$$

where r_n is the sensor noise variance. The *a posteriori* X_n is then updated using

$$X_n = X_n^- + K_n \left(I_n - H_n X_n^- \right) \tag{2.17}$$

where I_n is the raw frame. The *a posteriori* error covariance is then updated via,

$$P_n = (I_d - K_n H_n) P_n^-$$ (2.18)

where I_d is a 3 x 3 identity matrix. X_n contains three unknowns: amplitude, phase shift, and DC offset. A raw frame measurement is a one dimensional value, as such, at least three measurements are required to estimate phase using linear algebra. Constraints set in Q enable the Kalman filter to reduce the number of raw frames required to determine the phase shift (Streeter (2018)). The first constraint requires that Q be a diagonal matrix due to independence. The second constraint requires the Q component relating to the signal offset B to be set at a small nonzero value; based on the assumption that only small variations in B occur due to the differential measurement design of the pixels (Bamji *et al.* (2015); Streeter (2018)). When transverse motion occurs, the Kalman filter adapts to the change in phase within fewer raw frames. This results in a range image with fewer motion affected pixels, because the pixels that have observed a transverse motion event during earlier raw frames can compute a correct range. However, pixels that observe a transverse motion event immediately before this raw frame do not have sufficient raw frames to compute range correctly. A simulation of phase estimation using Kalman filter and running time-of-flight is shown in Fig. 2.12, which shows that at the fifth raw frame, the correct range is obtained using the Kalman filter method due to its ability to adapt to change more rapidly, while the running time-of-flight method cannot. It can also be seen that at the fourth raw frame, neither method has sufficient raw frames to obtain the correct range. In order to obtain a correct range computation for pixels with a recent transverse motion event, the Kalman filter is applied to the raw frames in reverse order. The raw frame which we want to compute a range for will, therefore, have two possible X_n values: one from the forward pass, and one from the reverse pass. The X_n with the lower prediction error is chosen, where the error is calculated as

$$e_n \ = \ |I_n - H_n X_n|.$$ (2.19)
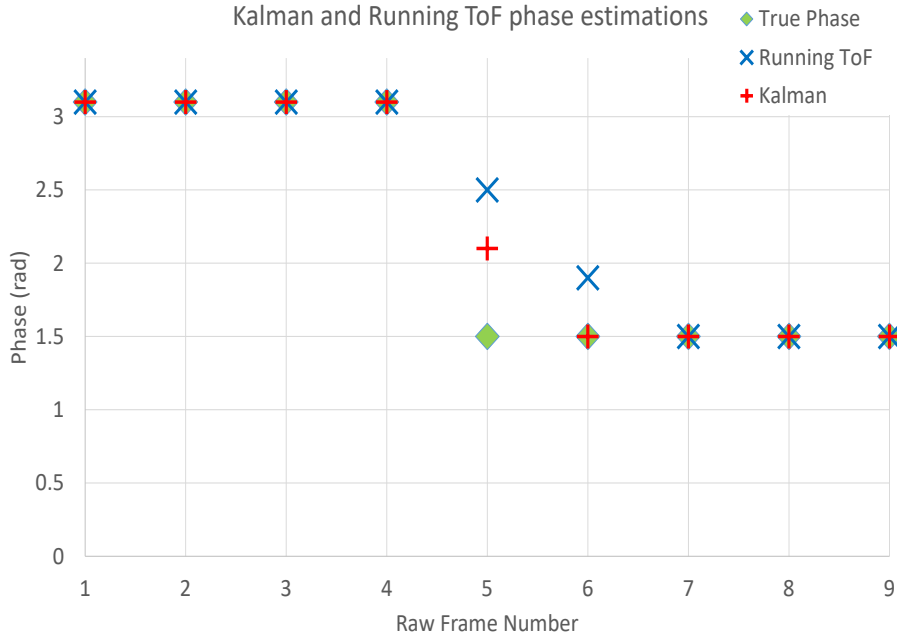
Figure 2.12: Simulation of phase estimation using Kalman and Running Time-of-Flight. The transverse motion event occurs between the fourth and fifth raw frames.

The flow chart in Fig. 2.13 summarises the time-of-flight Kalman filter run over the raw frames in a single direction.
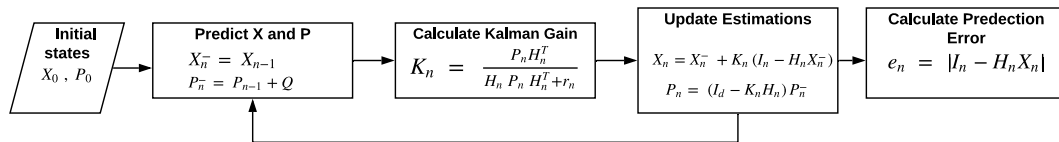
Figure 2.13: Time-of-Flight Kalman Filter Flowchart.

Running a Kalman filter on data forwards and backwards is referred to as a bidirectional Kalman filter (Hu *et al.* (2007); Goh *et al.* (2015)) Fig. 2.14 shows that although at the fifth raw frame, the forward Kalman is unable to obtain a correct phase estimation, the backward Kalman does, and will therefore be used when computing a range image.
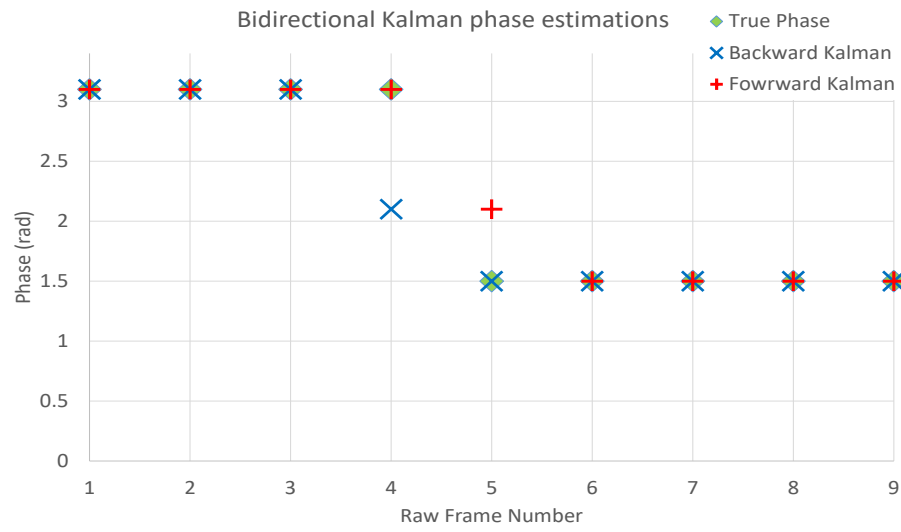
Figure 2.14: Simulation of bidirectional Kalman filter phase estimations. The transverse motion event occurs between the fourth and fifth raw frames.

Chapter 3

Adaptation of The Kalman Filter to Multi-Frequency Operation in Static Scenes

3.1 Introduction

The bidirectional Kalman filter described in chapter two operates on the assumption that if a scene is static, the best prediction of X_n is its previous value, because the amplitude, phase shift, and DC offset are not expected to change. Some commercial cameras operate over multiple modulation frequencies, enabling extended distance measurement, as well as detection and removal of data that may be corrupted by significant error. In multi-frequency operation, if the amplitude or phase shift changes between modulation frequencies in static scenes, then the assumption on which the time-of-flight Kalman filter depends is violated. In order to adapt the bidirectional Kalman filter to multi-frequency time-of-flight operation, such changes need to be investigated and taken into account. Since the bidirectional Kalman filter method relies on the Kalman filter running on the raw frames in forward and reverse order, in a multi-frequency operation, expected changes in the amplitude and phase shift need to be applied between modulation frequencies in both the forward

and backward Kalman filter operations.

The Microsoft Kinect V2 is a time-of-flight camera that uses three different modulation frequencies 80MHz, 16 MHz and 120 MHz. The camera captures nine raw frames in succession with three raw frames per modulation frequency at phase offsets 0, $2\pi/3$, and $4\pi/3$. As shown in Fig. 3.1, to obtain a range image corresponding to the fifth raw frame, expected changes in amplitude and phase are applied between the 80 MHz and the 16 MHz modulation frequencies for the forward Kalman filter, while for the backward Kalman filter, the changes in amplitude and phase are taken into account between 120 MHz and 16 MHz.
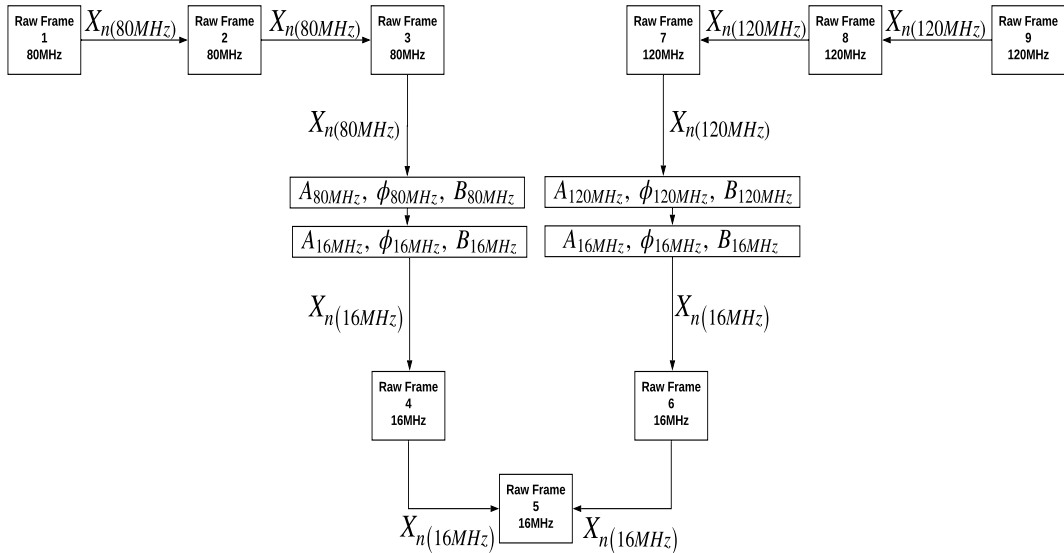
Figure 3.1: Adapted bidirectional Kalman filter flowchart

3.1.1 Amplitude

When capturing raw frames using different modulation frequencies of a static scene and calculating amplitudes from them, we find subtle differences in amplitudes that depend on the modulation frequency where higher modulation frequencies result in lower amplitude values. To demonstrate this, the Microsoft Kinect V2 is used to obtain raw frames using three modulation frequencies 80MHz, 16MHz and 120MHz. The DFT is applied to the the three sets to obtain an amplitude measurement for each frequency.

(a) 16 MHz amplitude image

(b) 80MHz amplitude image

(c) 120MHz amplitude image

Figure 3.2: Amplitudes obtained using different modulation frequencies

The differing amplitudes at different modulation frequencies result from the slew rate of the light sensor as well as the slew rate of the light source. As shown in Fig. 3.3, in order to observe the slew rate of the light source at different modulation frequencies and its impact on the light produced, the light source of a Microsoft Kinect V2 is placed in front of a collection lens. The coll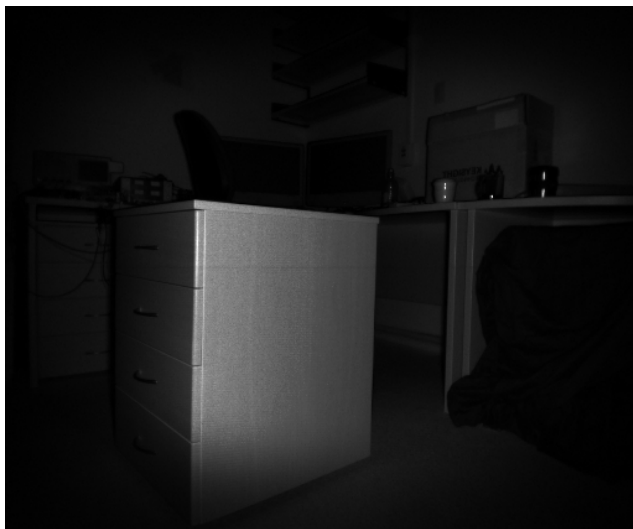ected light is measured by a photo detector, and the output voltage is measured by an oscilloscope. The result of the experiment shown in Fig. 3.4 shows that due to the slew rate, the proportion of the square wave at maximum amplitude changes depending on the modulation frequencies. At higher frequencies the proportion of the wave at maximum amplitude is lower. As such, the modulated light captured by the camera has high amplitudes at low modulation frequencies, and low amplitudes at high modulation frequencies. The change in amplitude due to the change in modulation frequency is expressed as

$$A_{f_{m2}} = k \ A_{f_{m1}}, \tag{3.1}$$

where, $A_{f_{m1}}$ and $A_{f_{m2}}$ are amplitudes obtained using two different modulation frequencies, and $k > 0$ is some constant dependent on the time-of-flight camera hardware.
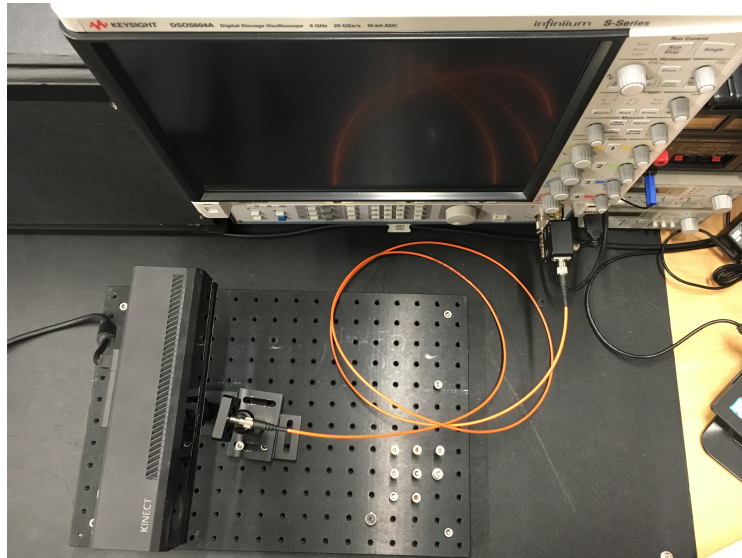
Figure 3.3: Slew rate of time-of-flight light source experiment
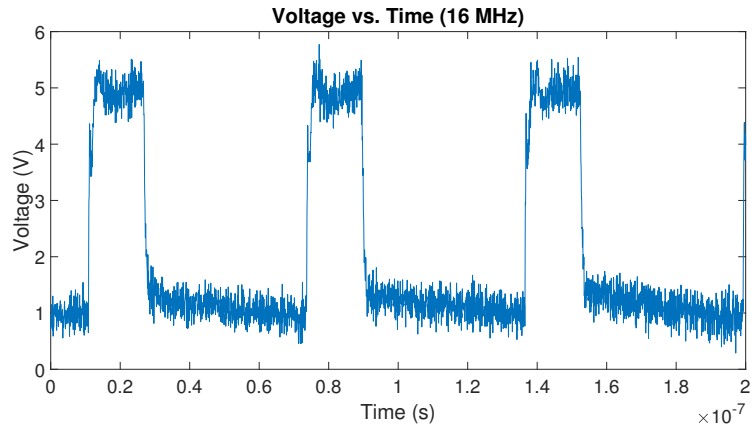
(a) Light at 16 MHz modulation frequency
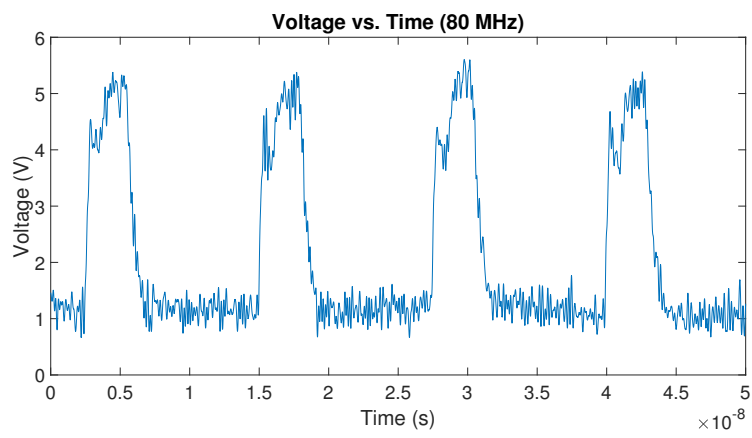
(b) Light at 80MHz modulation frequency
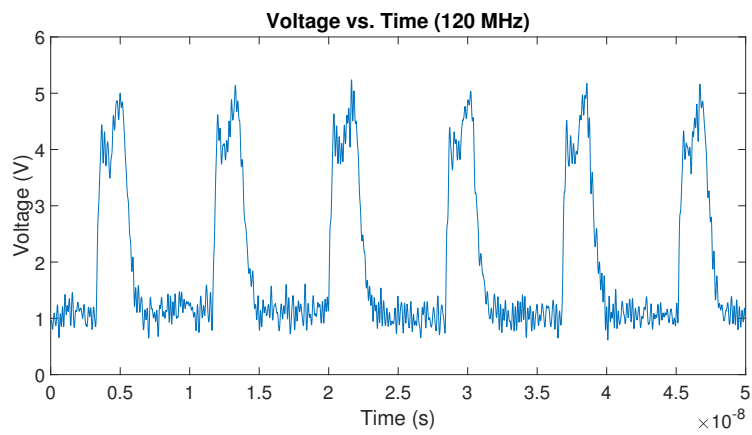
(c) Light at 120MHz modulation frequency

Figure 3.4: Light source slew rate experiment result

3.1.2   Phase Shift

The phase shift is also dependent on modulation frequency, as shown in Eqn. 2.2. As such, phase shifts obtained using different modulation have different values.
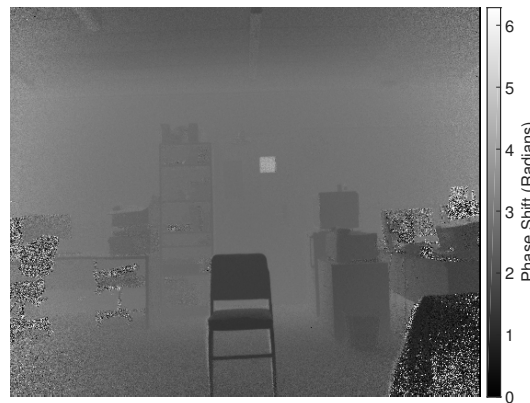
(a) 16MHz phase shift image
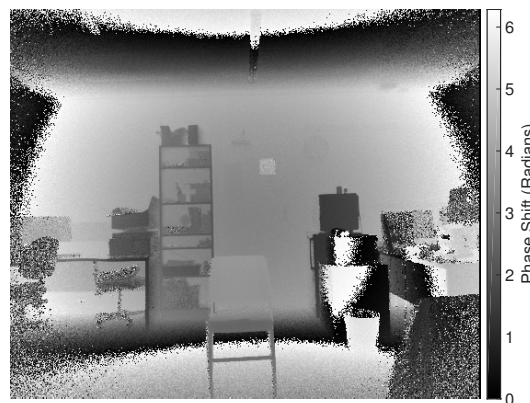
(b) 80MHz phase shift image
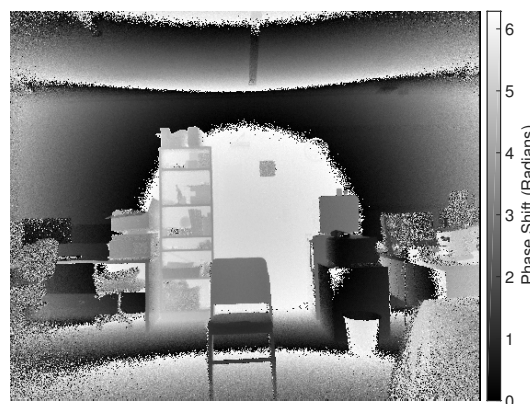
(c) 120 MHz phase shift image

Figure 3.5: Phase shifts obtained using different modulation frequencies

If the camera is operating on a scene that is within the maximum ambiguity range of the modulation frequency f_{m1}, we can predict the phase at modulation frequency f_{m2} using the equation

$$\phi_{f_{m2}} = \frac{f_{m2}}{f_{m1}} \left(\phi_{f_{m1}}\right) + s, \tag{3.2}$$

where $\phi_{f_{m1}}$ is the phase shift measured using modulation frequency f_{m1}, $\phi_{f_{m2}}$ is the predicted phase shift at modulation frequency f_{m2}, and s is an extra phase shift due to signal travel through the electronic hardware.

Eqn. 3.2 assumes that the camera is operating within the ambiguity range of the highest modulation frequency. In a scenario where this is not true, phases beyond 2π appear smaller due to phase wrapping. Operating the time-of-flight Kalman filter across multiple frequencies on phase wrapped data incurs non-linear error not accounted for by the Kalman filter model. Shifting the phase by an integer multiple of 2π removes this error, but creates a new ambiguity as the number of multiples of 2π must be accounted for, which is *a priori*-unknown. Predicting the phase shift across frequencies in a scene where phase wrapping could occur is performed according to

$$\phi_{f_{m2}} = \frac{f_{m2}}{f_{m1}} \left(\phi_{f_{m1}} + 2\pi N\right) + s, \tag{3.3}$$

where s is an extra phase shift due to signal travel through the electronic hardware, and the integer N represents the number of phase ambiguity intervals in $\phi_{f_{m1}}$. The multiplier N can be any integer from 0 to N_{max}, where,

$$N_{max} = \frac{f_1}{\text{GCF}\left(f_{m1}, \ f_{m2}\right)} - 1, \tag{3.4}$$

and GCF is the greatest common factor.

In multi-frequency time-of-flight Kalman filtering, X_n cannot be predicted to be the same as its previous value when the modulation frequency changes. When there is a change in modulation frequency X_n is recalculated with the

expected amplitude and phase shift.

$$
\begin{bmatrix}
kA\cos(\frac{f_{m2}}{f_{m1}}(\phi + 2\pi N) + s) \\
kA\sin(\frac{f_{m2}}{f_{m1}}(\phi + 2\pi N) + s) \\
B
\end{bmatrix}
\tag{3.5}
$$

X_n is calculated for every possible N value. This produces multiple X_n candidates. The X_n candidate that produces the smallest absolute residual in predicting the next two raw measurements is used as the *a priori* prediction.

$$
E_N = |I_n - H_n X_n| + |I_{n+1} - H_{n+1} X_n| \tag{3.6}
$$

The error for each candidate is calculated using two measurements because when a single measurement is used, multiple candidates can produce the same error.

For example, as shown in Fig. 3.6, an 80 MHz modulation frequency phase shift measurement of π results in 5 candidate phase shift values at 16 MHz. Fig. 3.6a shows that at π, two of these candidate phase shifts, at N = 1 and at N= 3, result in matching raw frame predictions at the zero phase step. However, as shown in Fig. 3.6b, the raw measurement prediction at the $\frac{2\pi}{3}$ phase offset does not result in matching measurements at N = 1 and at N=3. As such, using two raw measurement predictions to calculate the error for each candidate phase shift removes the ambiguity and narrows the candidates to one. The process of the adapted time-of-flight bidirectional Kalman filter described above is summarised by the flow chart in Fig. 3.7.
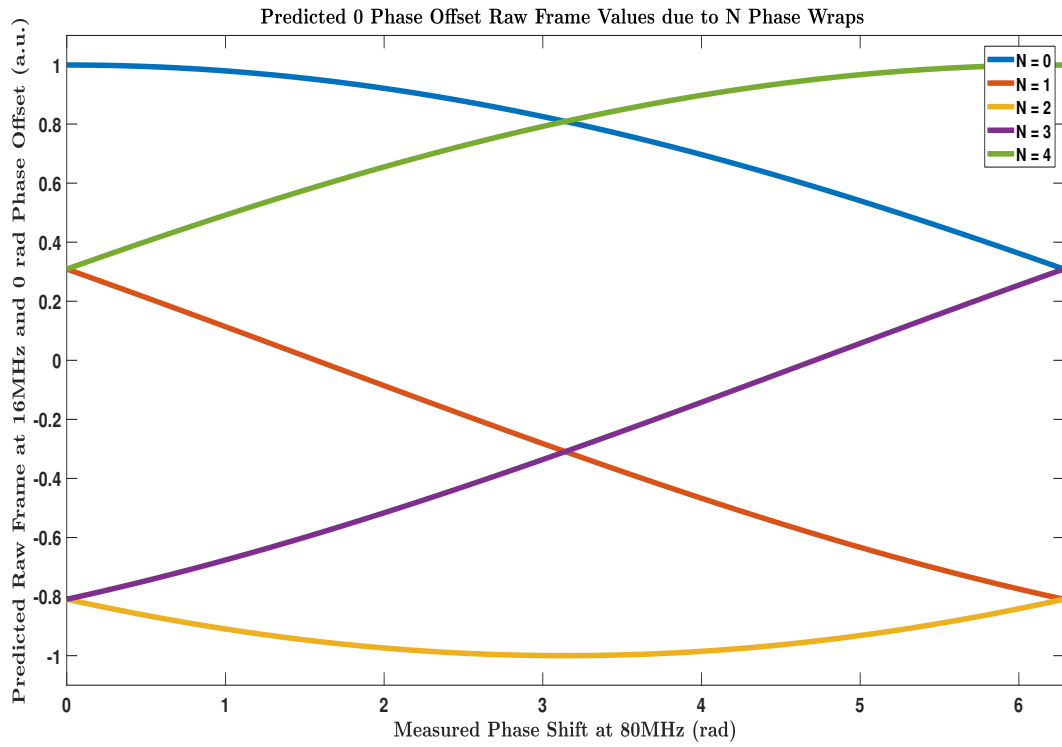
(a) Raw frame prediction at 0 phase offset
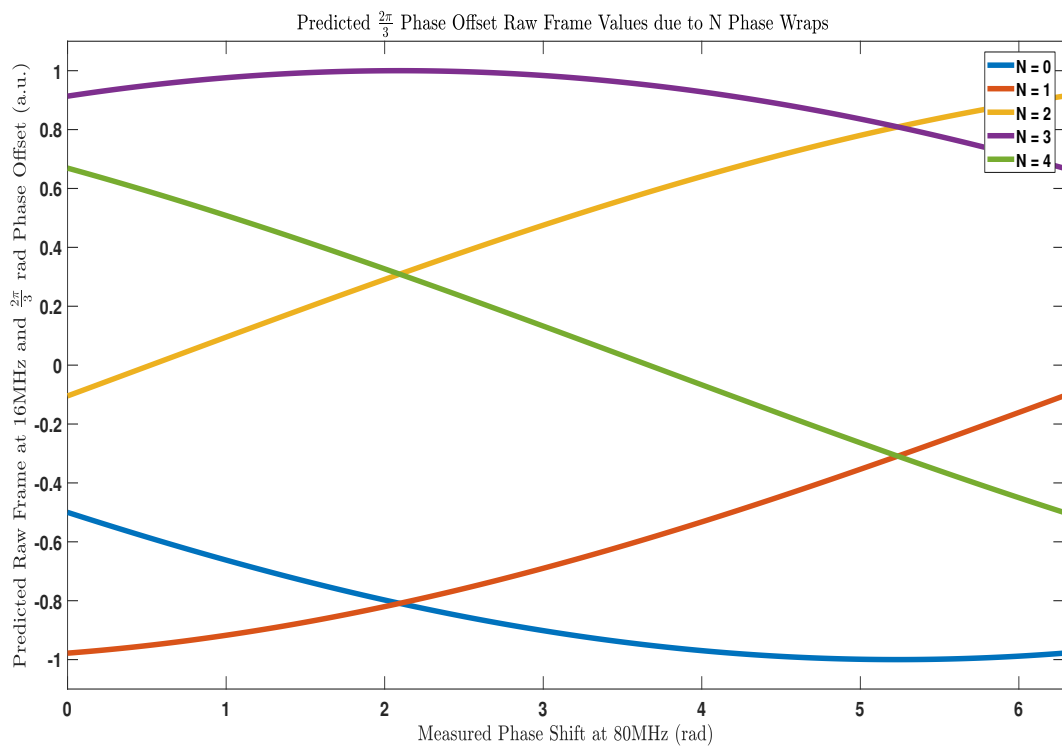
(b) Raw frame prediction at $\frac{2\pi}{3}$ phase offset

Figure 3.6: Raw frame predictions at two different phase offsets.
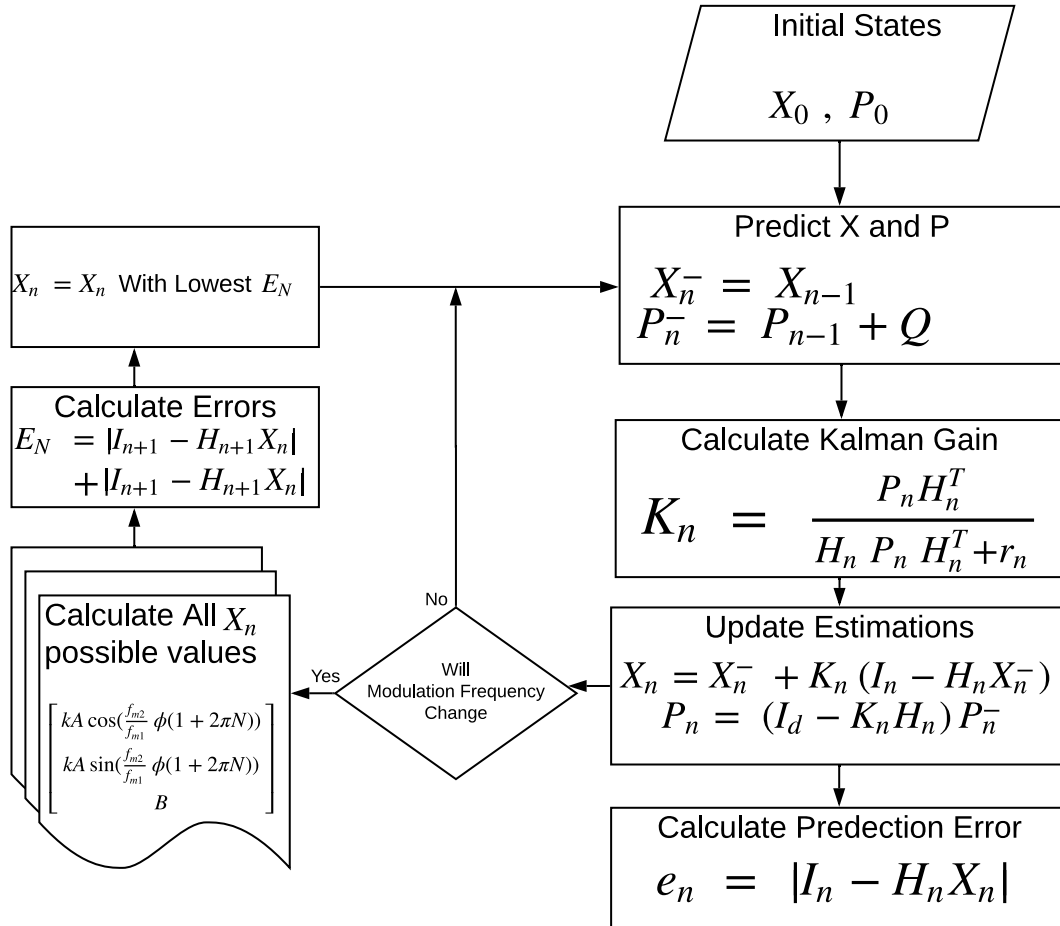
Figure 3.7: The adapted Kalman filter flow chart

3.2 Method

We test the performance of the adapted Kalman method using a Microsoft Kinect V2 time-of-flight camera. To obtain raw frames from the camera we use Libfreenect (Blake *et al.* (2016)), a user-space driver for the Microsoft Kinect. The CPU depth packet processor component within Libfreenect is edited to obtain raw frames before they are affected by any filters. All other data processing is performed using custom software written in Matlab R2017a and executed on a Windows 10 operating system. As stated previously, the Kinect uses three different modulation frequencies (80 MHz, 16 MHz and 120 MHz) and captures nine raw frames in succession with three raw frames per modulation frequency at phase offsets $\{0, 2\pi/3, 4\pi/3\}$.

3.2.1   Amplitude

To run the adapted Kalman filter method on the raw frames, the k parameter from Eqn. 3.1 must first be determined. This is done by capturing raw frame measurements of a static scene using the different modulation frequencies. The amplitude is calculated for each modulation frequency by performing the DFT over the corresponding raw frames. We solve for k for both the forward and backward Kalman frequency transitions by substituting the calculated amplitude in Eqn. 3.1.

We test the performance of the amplitude prediction across modulation frequencies by capturing amplitude images of a different static scene and using the 80MHz and 120MHz amplitude measurements to predict the 16 MHz amplitude measurement. We calculate the root mean squared error (RMSE) for the amplitude prediction using the calculated k values for both the forward and backward Kalman modulation frequency transitions and compare it to the RMSE of predicting the same amplitude (k=1).

3.2.2   Phase Shift

To run the adapted Kalman method, we also need to determine the s value in equation. 3.3. This is done by taking raw frame measurements of a static scene within the the maximum ambiguity range of the highest modulation frequency (120 MHz). We solve for s at each frequency transition by substituting the calculated phases in Eqn. 3.2. To test the performance of the phase prediction component of the adapted Kalman filter across modulation frequencies, we capture raw frames of a different scene, which is not constrained by the maximum ambiguity of 120MHz. The 16 MHz raw frames are used to calculate a reference phase image, while the 80MHz and 120 MHz raw frames are used as the preceding phase images. Eqn. 3.3 is used on the 80MHz and the 120 MHz phase images to predict all possible 16MHz phase candidates for the forward and backward frequency transitions, and Eqn. 3.6 is used to determine the predicted phase from the predicted candidates. We compare the phase predic-

tion using the proposed model to the phase prediction without the model by calculating the error in phase and the RMSE. The error in phase is calculated using

$$\phi_E = Angle\left(e^{i(\phi_R - \phi_P)}\right),\tag{3.7}$$

where ϕ_R is the reference phase, ϕ_P is the predicted phase, and Angle(z) is a function that returns phase angles for each element of complex array Z.

3.2.3 The Bidirectional Kalman Filter

For all the Kalman filter experiments, the Kalman filter parameters are set to the following:

$$P_0 = I_d, \quad r = 0.1, \quad Q = \begin{bmatrix} 0.5 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.01 \end{bmatrix}\tag{3.8}$$

We also perform an experiment to test the performance of the of adapted bidirectional Kalman filter and compare it to the performance of the un-adapted bidirectional Kalman method and the DFT method. To perform the unadapted bidirectional Kalman filter on a 16 MHz operation, nine raw frames at 16MHz are obtained. To create this, three sets of nine multi-frequency raw frames measurements are captured, the 16MHz raw frames within each measurement is used to simulate the 16 MHz data. The DFT is performed on 4th, 5th and 6th raw frames at 16 MHz to to create a reference phase image. We apply the unadapted bidirectional Kalman filter to both the multi-frequency operation and the simulated 16MHz single-frequency operation. The adapted Kalman filter method is performed on the multi-frequency data. The phase shift images corresponding to the 5th raw frame are calculated and compared, and Eqn. 3.7 is used to calculate the error and RMSE.

To investigate the effect of multi-path interference on the performance of the adapted bidirectional Kalman, we capture raw frames of a scene with expected multi-path interference. We then reduce the multi-path interference in the

same scene by covering the ground with a black sheet and capturing another set of raw frames. For both sets, we obtain a reference phase images for the 16MHz data using the DFT. we then use the adapted bidirectional Kalman filter to estimate the phase corresponding to the 5th raw frame. The error in phase, as well as the RMSE, is calculated for both predictions.

(a) Scene with multi-path   (b) Scene with reduced multi-path

Figure 3.8: Experiment performed to evaluate adapted bidirectional Kalman filter at varying multi-path interference

3.3   Result

3.3.1   Amplitude

The k values for both the forward Kalman filter (from 80MHz to 16MHz) and reverse Kalman fiter (from 120MHz to 16MHz) were calculated for the Microsoft Kinect by measuring amplitudes using the different modulation frequencies and solving for k. It was found that over the average of 217088 amplitude measurements, that $A_{f_{16}} = 1.24A_{f_{80}}$ and $A_{f_{16}} = 1.53A_{f_{120}}$. The accuracy of predictions is evaluated using a different static scene. The amplitude prediction from 80MHz to 16MHz is shown in Fig. 3.9, while the amplitude prediction from 120MHz to 16MHz is shown in Fig. 3.10

(a) Reference 16MHz amplitude
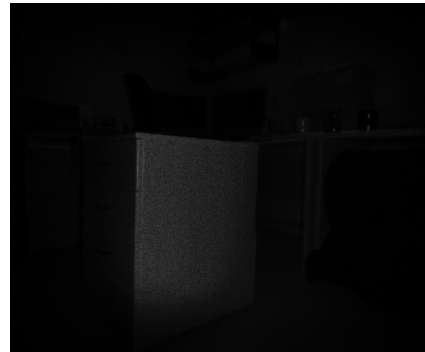
(b) Predicted amplitude

(without model)

(c) Predicted amplitude error

(without model)
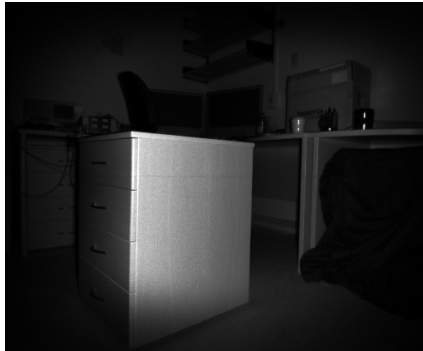
(d) Predicted amplitude

(with model)

(e) Predicted amplitude error

(with model)

Figure 3.9: The figures are obtained to evaluate the amplitude prediction from 80MHz to 16MHz. (a) is the 16MHz reference amplitude image. (b) is the amplitude calculated using the DFT of the 80MHz data. (d) is the 16MHZ amplitude predicted from the 80MHz using k = 1.24. (c) and (e) are the absolute error images of (b) and (d), calculated using the reference image.
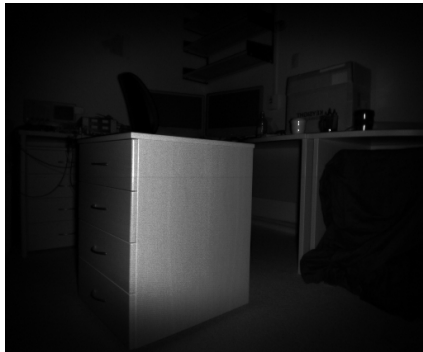
(a) Reference 16MHz amplitude

(b) Predicted amplitude
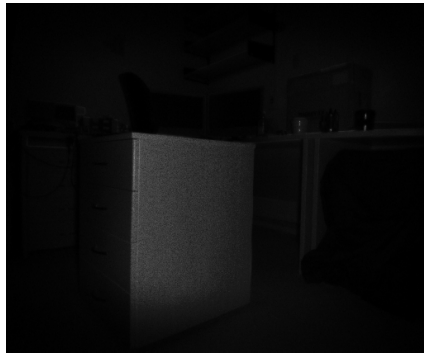
(without model)

(c) Predicted amplitude error

(without model)

(d) Predicted amplitude
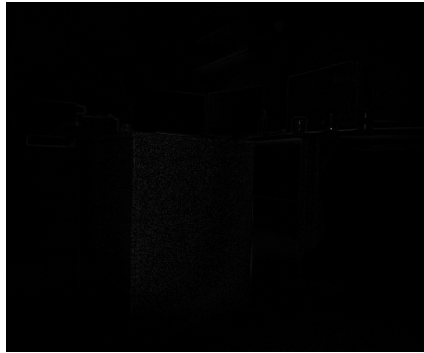
(with model)

(e) Predicted amplitude error

(with model)

Figure 3.10: The figures are obtained to evaluate the amplitude prediction from 120MHz to 16MHz. (a) is the 16MHz reference amplitude image. (b) is the amplitude calculated using the DFT of the 120MHz data. (d) is the 16MHZ amplitude predicted from the 120MHz using k = 1.53. (c) and (e) are the absolute error images of (b) and (d) calculated using the reference image.

Table 3.1: RMSE of amplitude predictions with and without model

|	Amplitude RMSE without Model	Amplitude RMSE with Model
80 MHz to 16 MHz	**0.069**	**0.023**
120 MHz to 16 MHz	**0.095**	**0.025**

The reduction in error seen in Fig. 3.9 and Fig. 3.10, as well as the reduction in RMSE shown in Table. 3.1, shows that amplitude can be predicted more accurately using the model. In Fig, 3.9e and Fig, 3.10e, the error in amplitude appears more significant in high amplitude regions of the scene.

However, the error relative to the actual amplitude is more indicative of how the amplitude prediction will effect the performance of the time-of-flight Kalman Filter. The equation below is therefore used to calculate the relative amplitude errors.

$$Relative\ Error\ =\ \frac{|A_p - A_m|}{\max{(A_p,\ A_m)}} \tag{3.9}$$

The relative error images in Fig. 3.11 indicate that the accuracy of the amplitude prediction is highly dependent on the magnitude of the signal received. This is attributed to the signal to noise ratio, where the lower the signal received, the larger the impact noise has on the prediction. The result, however, indicates that the adapted Kalman's filter performance is expected to be lower at low amplitudes.
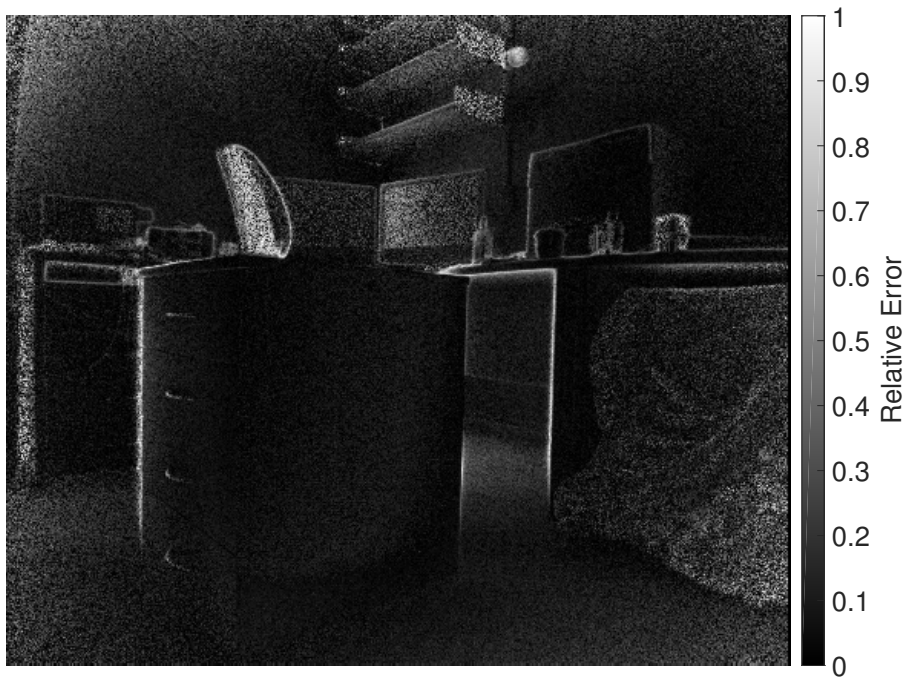
(a) 80MHz relative amplitude error
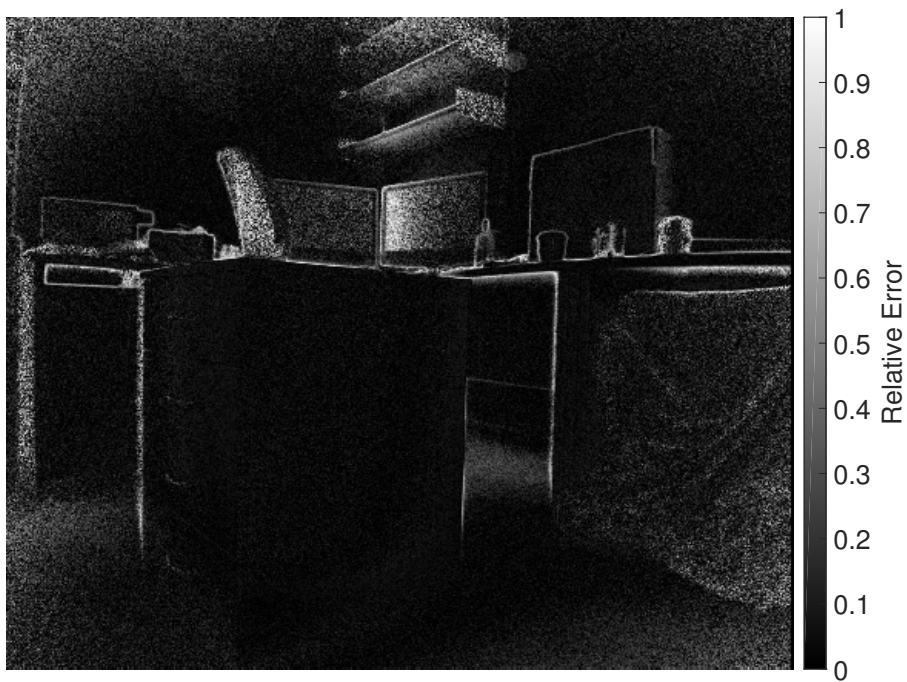
(b) 120MHz relative amplitude error

Figure 3.11: Relative error of amplitude predictions. (a) is the relative error of the predicted 16MHz amplitude using the calculated 80MHz amplitude. (b) is the relative error of the predicted 16MHz amplitude using the calculated 120MHz amplitude.

3.3.2 Phase Shift

Phase shift prediction between modulation frequencies using Eqn. 3.3 requires the s parameters to be determined. The s parameters for the 80MHz to 16MHz phase shift prediction and the 120MHz to 16MHz phase shift prediction were calculated by measuring the phase of a static scene using the three modulation frequencies and solving for s using Eqn. 3.2. As shown in Table. 3.2, the s parameters are found to be dependent on the location of the pixel in the camera.

Table 3.2: Parameter result for the phase prediction across modulation frequencies. Odd and even rows of camera pixels are represented by (Odd) and (Even).

	80MHz To 16MHz	120MHz To 16MHz
s (Odd)	0.75	0.3
s (Even)	0.13	0.9

We test the accuracy of the phase prediction using Eqn. 3.3 with the s values above. We capture phase images of a static scene and use the 80MHz and 120MHz phase measurements to predict the 16MHz phase measurement. The phase shift prediction from 80MHz to 16MHz is shown in Fig. 3.12, while the phase shift prediction from 120MHz to 16MHz is shown in Fig. 3.13. The reduction in error seen in Fig. 3.12 and Fig. 3.13, as well as the reduction in RMSE shown in Table. 3.3, show that we can predict the phase more accurately using the model.

(a) Reference 16MHz phase shift
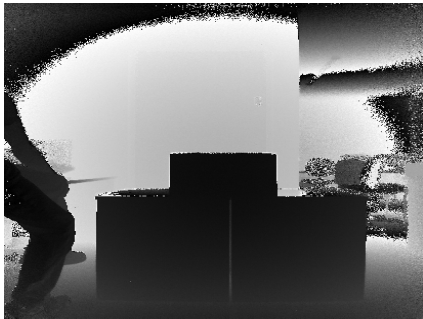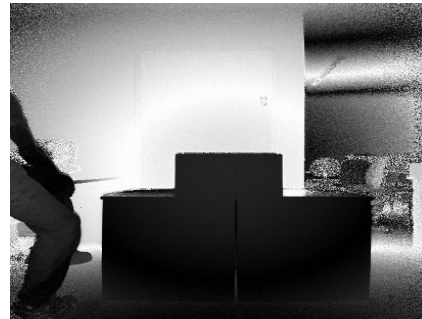
(b) Predicted phase shift

(without model)

(c) Predicted phase error

(without model)
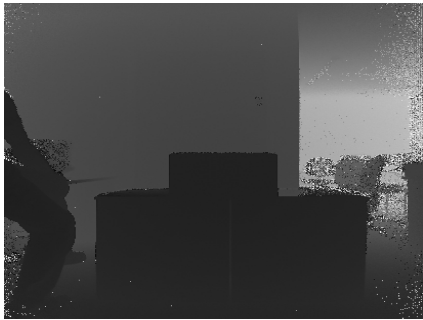
(d) Predicted phase shift
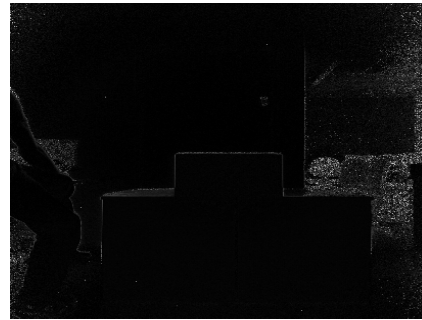
(with model)

(e) Predicted phase error

(with model)

Figure 3.12: The figures are obtained to evaluate the phase prediction from 80MHz to 16MHz. (a) is the 16MHz reference phase image. (b) is the phase predicted directly from the DFT of the 80MHz data. (d) is the 16MHz phase predicted from the 80MHz using Eqn. 3.3. (c) and (e) are the absolute error images of (b) and (d) calculated using the reference image.

(a) Reference 16MHz phase shift
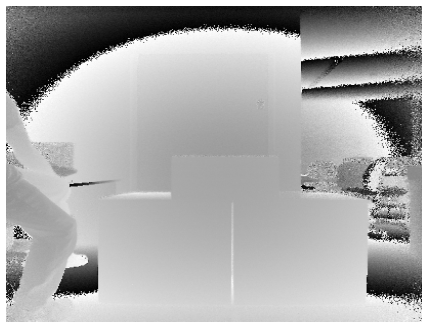
(b) Predicted phase shift
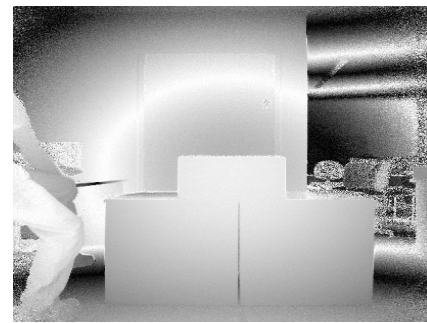
(without model)

(c) Predicted phase error

(without model)

(d) Predicted phase shift

(with model)

(e) Predicted phase error

(with model)

Figure 3.13: The figures are obtained to evaluate the phase prediction from 120MHz to 16MHz. (a) is the 16MHz reference phase image. (b) is the phase predicted directly from the DFT of the 120MHz data. (d) is the 16MHz phase predicted from the 120MHz using Eqn. 3.3. (c) and (e) are the absolute error images of (b) and (d) calculated using the reference image.

Table 3.3: RMSE of phase predictions with and without model

	Phase RMSE without Model	Phase RMSE With Model
80MHz to 16MHz	1.775	0.256
120MHz to 16MHz	2.244	0.269

3.3.3   Bidirectional Kalman Filter

The phase estimations using the unadapted bidirectional Kalman and the adapted bidirectional Kalman are compared using the FFT reference image.

Table 3.4: RMSE for adapted and unadapted bidirectional Kalman for single and multi-frequency operation in non dynamic scene.

	RMSE (Multi-Frequency) Operation	RMSE (Single-Frequency) Operation
Unadapted Bidirectional Kalman Filter	0.351	0.02
Adapted Bidirectional Kalman Filter	0.23	NA

The error images in Fig. 3.14, as well as the RMSE shown in Table. 3.4, show that in multi-frequency operation, the adapted bidirectional Kalman filter performs better than the unadapted bidirectional Kalman filter. The adapted Kalman filter in multi-frequency operation, however, does not match the accuracy of the unadapted bidirectional Kalman in a single modulation frequency operation. This is because in a non dynamic scene, the accuracy of the unadapted bidirectional Kalman's phase prediction is only dependent

on the noise, whilst the accuracy of the prediction of the adapted Kalman is highly dependent on the model's accuracy between modulation frequencies, as well as noise. The adapted Kalman filter model cannot perfectly predict the phase because the model does not take in to account multi-path error, and because multi-path error is modulation frequency dependent. If an increase or decrease occurs in the phase measurement due to multi-path interference, the adapted Kalman fails to take this into account in the prediction. The images in Fig. 3.15 are of a scene with a large amount of multi-path interference. Once the a black sheet is used to cover the floor, multi-path error is reduced by removing the reflection of the floor. As Table. 3.5 shows, the accuracy of the adapted bidirectional Kalman improves in low multi-path scenes.

Table 3.5: RMSE of adapted Kalman filter under high and low multi-path interference

	RMSE
Scene With High MPI	0.168
Scene With Low MPI	0.148

(a) Unadapted Kalman phase
(multi-frequency)
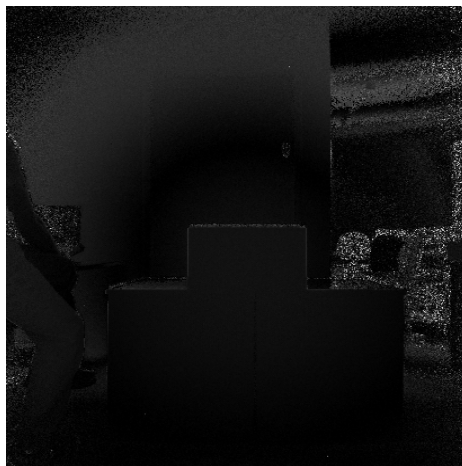
(b) Unadapted Kalman phase error
(multi-frequency)

(c) Adapted Kalman phase
(multi-frequency)

(d) Adapted Kalman phase Error
(multi-frequency)

(e) Kalman phase estimation from
single frequency

(f) Kalman phase estimation error
from single frequency

Figure 3.14: Comparison of phase estimation between the adapted and un-adapted bidirectional Kalman filters in non dynamic scene.

(a) High multi-path scene
reference phase
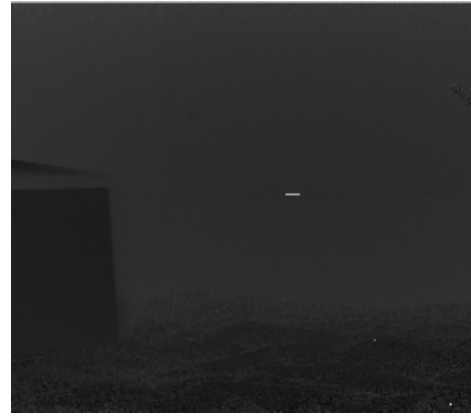
(b) High multi-path scene
reference phase
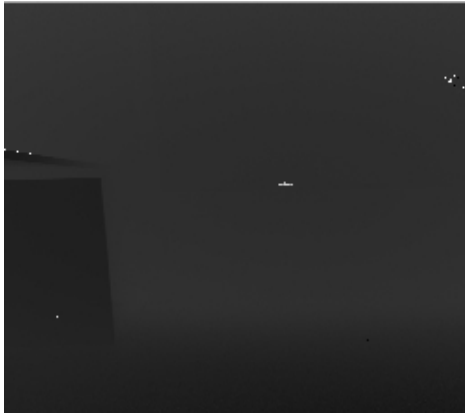
(c) High multi-path scene
phase estimation

(d) Low multi-path scene
reference phase
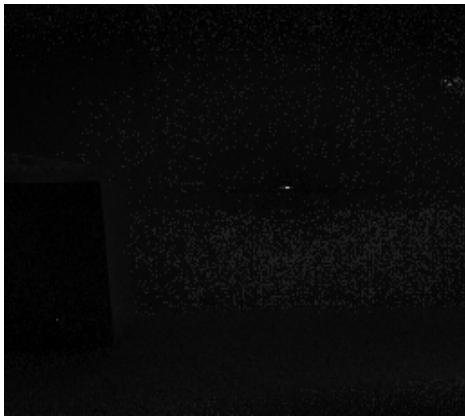
(e) High multipath scene Phase
estimation error

(f) Low multi-path scene phase
estimation error

Figure 3.15: Comparison of phase estimation by adapted bidirectional Kalman
filter in varying multi-path scenes

Chapter 4

Motion Correction Using Adapted Bidirectional Kalman Filter

4.1 Introduction

The purpose of the time-of-flight bidirectional Kalman filter is to reduce the error caused by transverse motion. Although adapting the bidirectional Kalman filter in non dynamic scenes to provide accurate phase estimations is necessary, the adapted Kalman filter in multi-frequency operation remains ineffectual if it cannot fulfill its purpose of reducing motion error in dynamic scenes. In this Chapter, the adapted Kalman filter proposed in the previous chapter is tested on dynamic scenes, and changes are made to the model to enable motion error reduction.

When X_n is initialised, The Kalman filter needs at least three consistent and consecutive raw frames to reach an accurate phase estimation. As such, when a transverse motion event occurs between the initial three raw frames, the phase estimated at the third raw frame by the Kalman filter becomes incorrect. In single modulation frequency operation, as shown in the diagram in Fig. 4.1, the unadapted bidirectional Kalman filter is still able to obtain

a correct composite phase estimation by using the phase estimated by the Kalman filter in the reverse direction. The bidirectional Kalman filter selects the composite phase from the forward pass or the reverse pass depending on how close they can predict the corresponding raw frame using Eqn. 2.19.
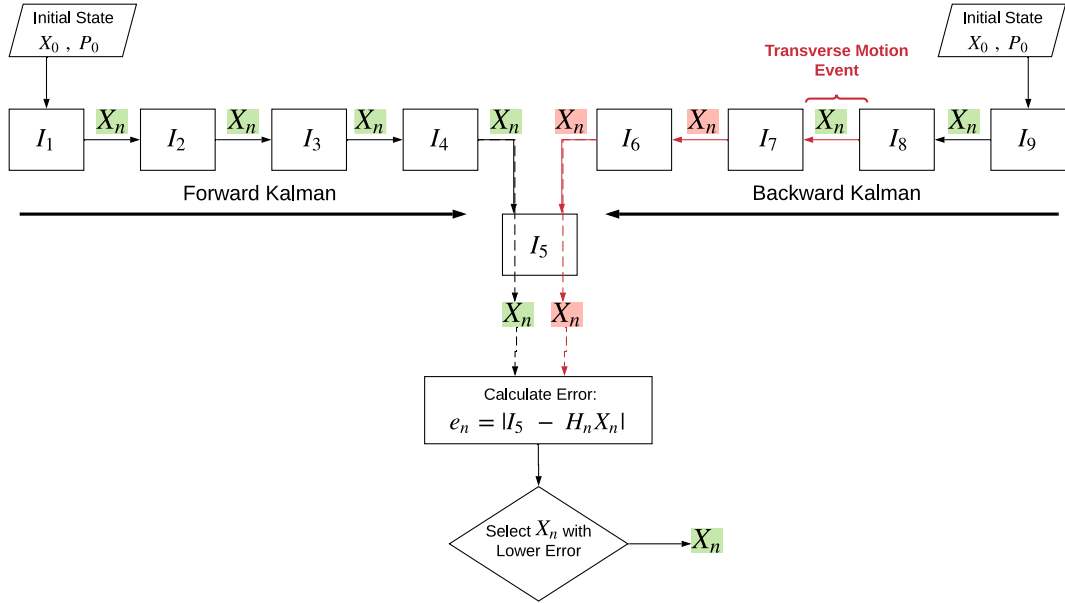
Figure 4.1: Unadapted bidirectional Kalman filter in single frequency operation flowchart. I_n represents the N raw frame. X_n highlighted in red represents an incorrect X_n estimation due to a transverse motion event, and X_n highlighted in green represents a correct X_n estimation.

4.1.1 Further Adaptations of the Adapted Bidirectional Kalman Filter

In multi-frequency operation, the adapted Kalman filter also fails to estimate the correct phase when a transverse motion event occurs within the first three raw frames after a zero initialisation of X_n. As in the unadapted Kalman filter, a correct phase estimation is obtained for the corresponding raw frame using the reverse pass. However, the phase shift prediction across modulation frequencies causes an incorrect selection of the composite phase using Eqn. 2.19.

When an incorrect phase is estimated due to a transverse motion event, the unwrapping component of the phase prediction method chooses the phase

shift candidate that best predicts the next two raw frames using Eqn. 3.6. If the number of possible phase candidates calculated using Eqn. 3.4 is high, the incorrect phase prediction results in an accurate raw frame prediction even if the predicted phase shift is incorrect. The unwrapping methodology used thus guarantees a low error. As such, Eqn. 2.19 fails to identify the correct phase estimation and leads to an incorrect composite phase selection.

To solve this problem, Eqn. 4.1 below is used to determine the composite phase shift instead of Eqn. 2.19. This is because the error calculated using Eqn. 4.1 relies on the prediction of raw frames that are not used by the phase unwrapping method. Fig. 4.2 shows that the use of Eqn. 4.1 rather than Eqn. 2.19, results in a correct composite phase selection in multi-frequency operation.

$$E_N \;=\; 2\left|I_{n-1} - H_{n-1}X_{n-1}\right| + 6\left|I_n - H_nX_n\right| \;+\; 2\left|I_{n+1} - H_{n+1}X_{n+1}\right| \quad (4.1)$$
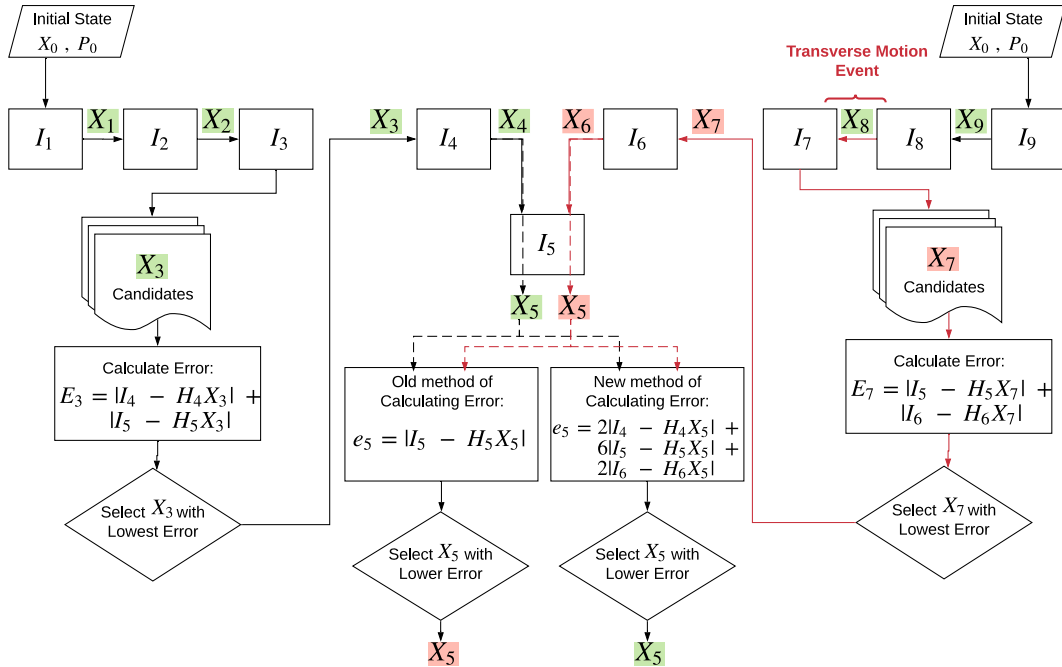

Figure 4.2: Flowchart of adapted bidirectional Kalman filter in multi-frequency operation. The flow chart shows two methods of calculating error for the composite X_n. The flow chart indicates that the new method selects the correct X_n

4.2   Method

In this section, we perform two quantitative experiments and one qualitative experiment.

The first quantitative experiment is used to compare the composite phase shift determined using the suggested method using Eqn. 4.1 to the composite phase shift determined using Eqn. 2.19. The experiment tests whether, the suggested method can select the correct phase shift, where the previous method is expected to fail. We perform the quantitative experiment by placing a board 1.3m away from the camera, as shown in Fig. 4.3. raw frames are captured of the board at two different positions with 20 cm between the two positions to mimic transverse motion. An artificial set containing nine raw frames is created to simulate data from multi-frequency operation. To simulate motion that occurs only between the 8th and the 9th raw frames, the first eight raw frames are created from the board at the first position, while the 9th raw frame is created from the measurement of the board at the second position. The DFT is applied to the 16 MHz measurements taken at the first position of the board to create a reference phase shift that is not affected by motion. We perform the adapted bidirectional Kalman filter method on the data, and create a composite phase image using Eqn. 4.1 and a composite phase image using Eqn. 2.19 corresponding to the 5th raw frame. The error in phase is calculated for both methods by using the reference image and Eqn. 3.7.

The second quantitative experiment is conducted to test the performance of the adapted bidirectional Kalman method in dynamic scenes by comparing it to the unadapted bidirectional Kalman filter method and the DFT method. The experiment is performed by placing a board 1.3m away from the camera. Raw frames are captured of the board at nine different positions with 5cm between each position to mimic continuous transverse motion. The DFT is applied to the 16 MHz measurements taken at the 5th position to create a reference phase image that is not affected by motion. Several artificial sets that simulate the motion of the board are created using the raw measurements.

Figure 4.3: Quantitative experiment setup

The first set mimics the camera operation, containing nine raw frames with modulation frequencies of 80MHz, 16MHz and 120MHz. For this set, phase images corresponding to the 5th raw frame are calculated using the adapted bidirectional Kalman filter method, the unadapted bidirectional Kalman filter method, and the DFT method. The phase error is calculated for each method using the reference phase image. The process is repeated 5700 times using pixels from regions with motion, and the RMSE is calculated. To compare the performance of the adapted and unadapted bidirectional Kalman filter methods, the percentage of instances with less error than the DFT method is calculated. We also compare the performance of the unadapted bidirectional Kalman filter method operating on a single modulation frequency to the performance of the adapted bidirectional Kalman filter method operating on raw frames with multiple modulation frequencies. We achieve this by creating a second set consisting of nine raw frames that simulate continuous motion at a modulation frequency of 16 MHz. Phase images corresponding to the 5th raw frame are calculated using the unadapted bidirectional Kalman filter method. The reference phase image is used to calculate the RMSE and the percentage of instances the unadapted bidirectional Kalman filter method performed better than the DFT method. To further test the proposed model, the quantitative

Table 4.1: List of tested foreground and background distances.

Position	1	2	3	4	5	6
Background distance (m)	1.71	2.61	2.88	3.39	3.78	4.12
Foreground distance (m)	1.19	1.67	0.88	0.96	2.18	1.17

experiment is repeated at varying foreground and background distances shown in table. 4.1. The distances are controlled by repositioning both the camera and the board.

The qualitative experiment is carried out to test the performance of the adapted bidirectional Kalman method in dynamic scenes containing both intra-frame motion and inter-frame motion by comparing it to the performance of the unadapted bidirectional Kalman method and of the DFT method. To perform the qualitative experiment, nine raw frames are captured of a board that is swung in a transverse fashion approximately 1m away from the camera. A phase image corresponding to the 5th raw frame is calculated using the adapted bidirectional Kalman filter method, the unadapted bidirectional Kalman filter method, and the DFT method. To obtain the phase using the DFT method, raw frames 4 to 6 are used. To obtain phase using the bidirectional Kalman filter method, we use raw frames 1 to 6 for the forward pass and raw frames 4 to 9 for the reverse pass. Eqn. 4.1 is used to determine the composite phase at the 5th raw frame.

4.3 Results

The first quantitative experiment is performed to evaluate the proposed composite phase selection method, which was done by introducing motion between the 8th and 9th raw frames, and testing both composite phase shift selection methods at the 5th raw frame. As shown in Fig. 4.4, the phase shift images, as well as the phase shift error images, show that the phase shift estimated using the proposed method produces a more accurate phase shift estimation.

(a) True phase shift
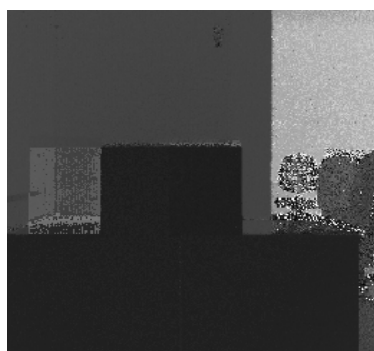
(b) Phase shift using
previous method
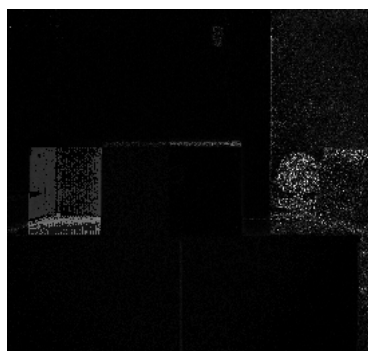
(c) Phase error using
previous method
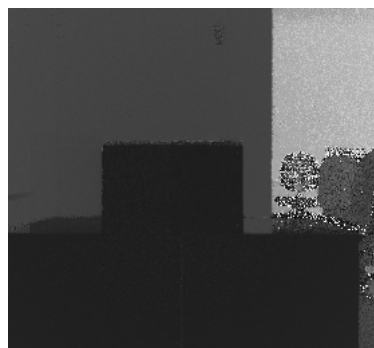
(d) Phase shift using
proposed method
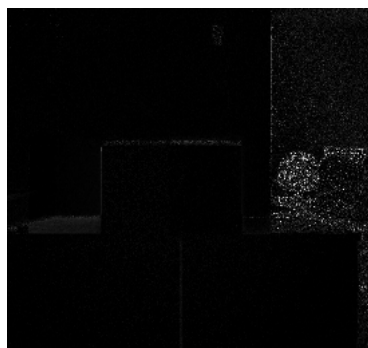
(e) Phase error using
proposed method

Figure 4.4: The previous method refers to the composite phase selection determined using a single raw frame prediction using Eqn. 2.19. The proposed method refers to the composite phase selection determined using three raw frame predictions using Eqn. 4.1.

A more in-depth analysis of the result is performed by plotting the phase shift, amplitude, and raw frame estimates based on the X_n produced by the adapted bidirectional Kalman filter at every raw frame. As expected, although the phase shift estimated by the backward Kalman filter is incorrect at the 5th raw frame, shown in Fig. 4.5a, its corresponding raw frame prediction shown in Fig. 4.5b appears accurate. This is due to the fact that the unwrapping of the phase shift between modulation frequencie of 120MHz and 16MHz produces 52 X_n candidates according to Eqn. 3.4. When the raw frame predictions at the 5th raw frame alone are used to select a composite, an incorrect selection occurs. However, taking the 4th, 5th and 6th raw frame predictions into account leads to an accurate composite phase selection as shown in Fig. 4.6a

In the second quantitative experiment, where transverse motion occurs continuously between frames, we test the performance of the proposed adapted bidirectional Kalman filter and compare it to the unadapted bidirectional Kalman filter, as well as the traditionally used DFT method. In the experiment, the phase shift image obtained using the DFT, as shown in Fig. 4.9, incurs significant error to the left and right of the board, where the transition between foreground and background occurs. The phase image obtained using the unadapted bidirectional Kalman filter shows an increase in error when compared to the DFT. The increase in error is due to its incompatibility with the multi-modulation operation, while the DFT range image is obtained using raw frames belonging to a single modulation frequency. When the proposed adapted bidirectional Kalman filter is used on the multi-frequency operation, the error in the region of motion is significantly reduced. Some error remains at the edges of the board; we believe the remaining error is due to mixed pixels. It is also observed that errors increase in regions where multi-path is expected; as discussed in Chapter three, these errors are due to the fact that multi-path error is modulation frequency-dependent, and the model uses Eqn. 3.3 which does not take multi-path into account when predicting phases.
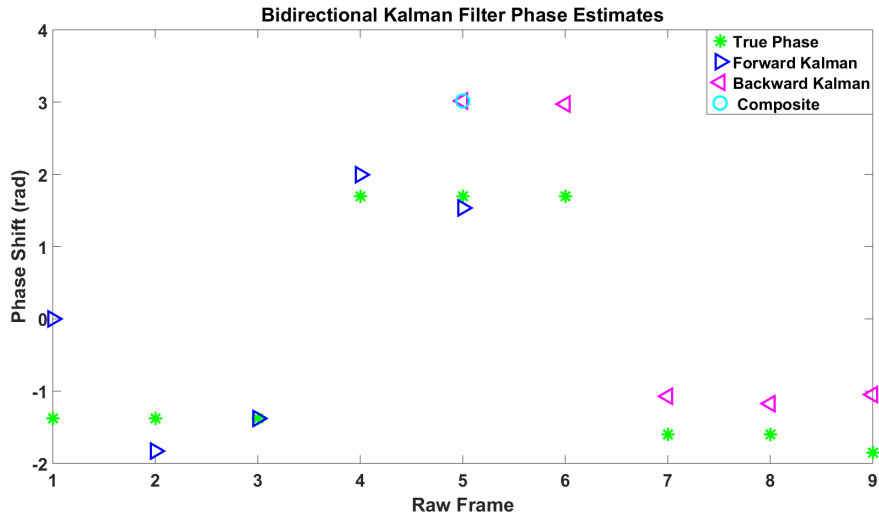
(a) Phase shift estimations using adapted bidirectional Kalman filter.
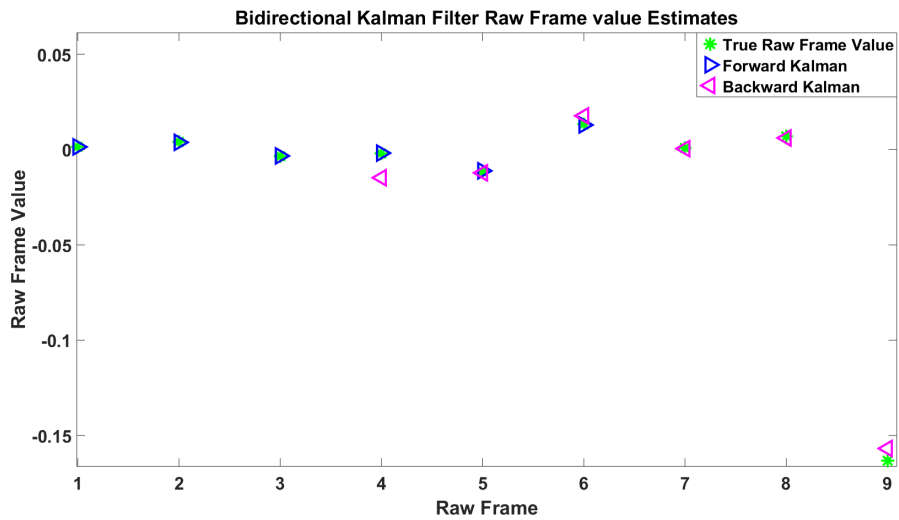
(b) Raw frame estimations using adapted bidirectional Kalman filter.

Figure 4.5: The charts are of phase shift, and raw frames estimates based on the X_n produced by the adapted bidirectional Kalman filter at every raw frame. The composite phase shift at the 5th raw frame is selected using a single raw frame using Eqn. 3.6.

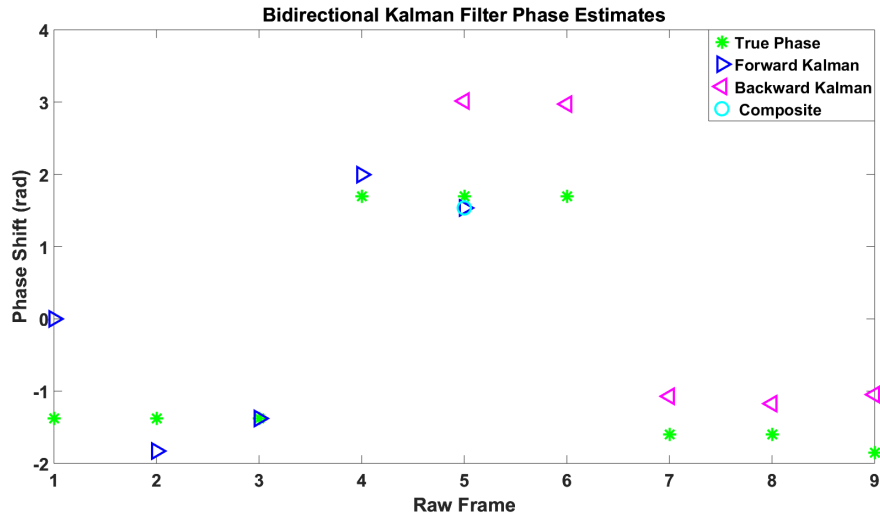(a) Raw frame estimations using adapted bidirectional Kalman filter.
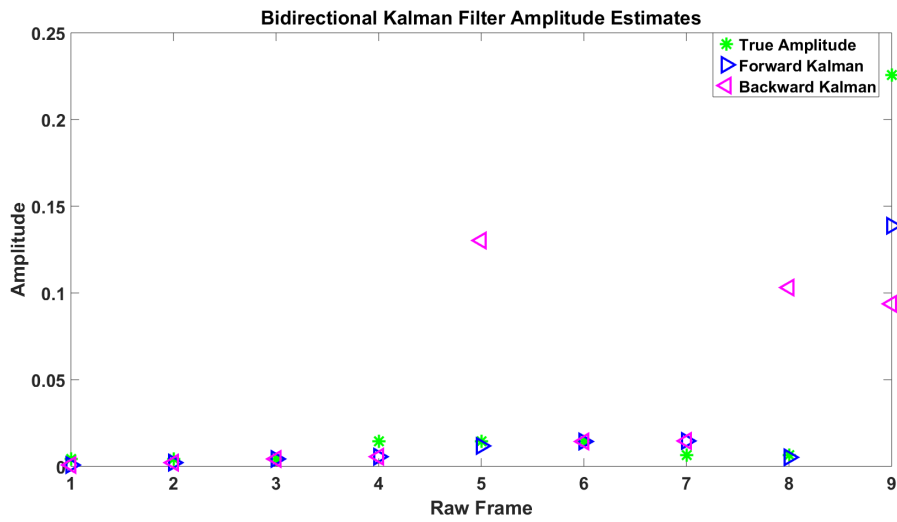
(b) Amplitude estimations using adapted bidirectional Kalman filter.

Figure 4.6: The charts are of result phase shift and amplitude estimates based on the X_n produced by the adapted bidirectional Kalman filter at every raw frame. The composite phase shift at the 5th raw frame is selected using three raw frames using Eqn. 4.1
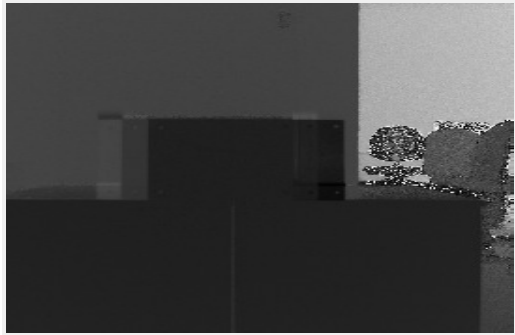
(a) DFT phase image
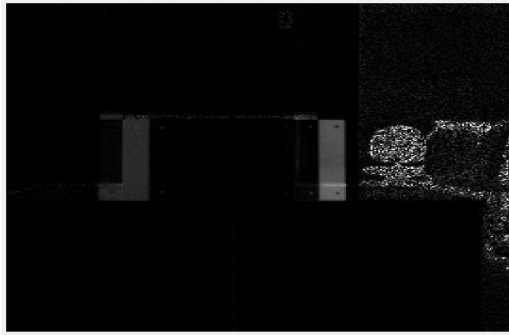
(b) DFT error image
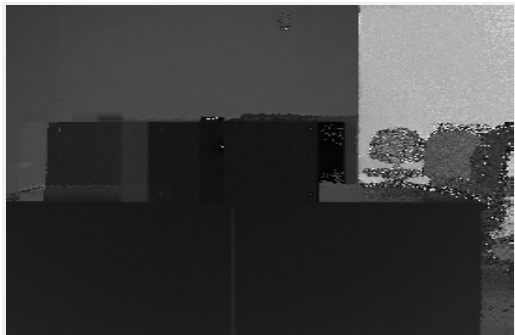
(c) Unadapted bidirectional

Kalman filter phase
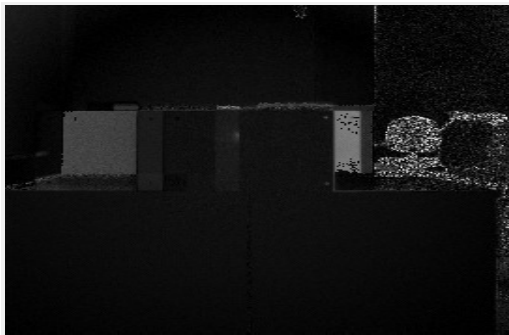
(d) Unadapted bidirectional

Kalman filter error
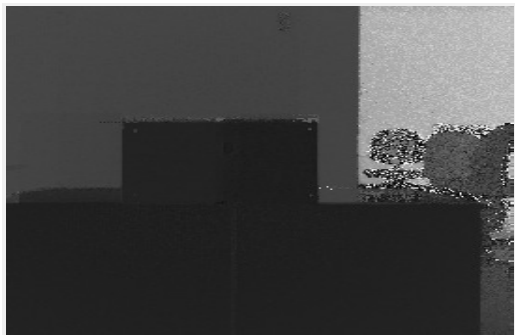
(e) Adapted bidirectional

Kalman filter phase
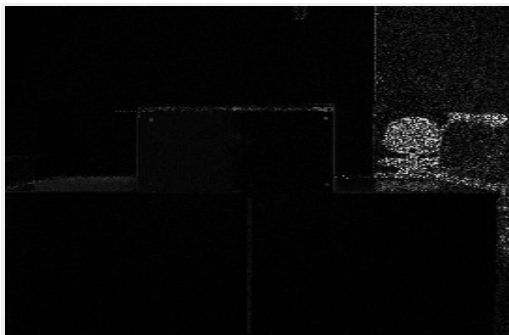
(f) Adapted bidirectional

Kalman filter error

Figure 4.7: Comparison of phase and error images recovered from the multifrequency quantitative experiment. (a) and (b) are obtained using the traditional DFT method. (c) and (d) are obtained using unadapted bidirectional Kalman filter method. (e) and (f) are from the proposed adapted bidirectional Kalman filter method.

(a) DFT method position(1)

(b) Proposed method position(1)

(c) Unadapted method position(1)

(d) DFT method position(2)

(e) Proposed method position(2)

(f) Unadapted method position(2)

(g) DFT method position(3)

(h) Proposed method position(3)

(i) Unadapted method position(3)

Figure 4.8: Quantitative experiment error images at positions 1, 2 and 3. The positions refer to specified foreground and background distances in Table. 4.1. (a), (d) and (g) are obtained using the traditional DFT method. (b), (e) and (h) are obtained using the proposed adapted bidirectional Kalman filter method. (c), (f) and (i) are obtained using the unadapted bidirectional Kalman filter method.

(a) DFT method position(4)

(b) Proposed method position(4)

(c) Unadapted method position(4)

(d) DFT method position(5)

(e) Proposed method position(5)

(f) Unadapted method position(5)

(g) DFT method position(6)

(h) Proposed method position(6)

(i) Unadapted method position(6)

Figure 4.9: Quantitative experiment error images at positions 4, 5 and 6. The positions refer to specified foreground and background distances in Table. 4.1. (a), (d) and (g) are obtained using the traditional DFT method. (b), (e) and (h) are obtained using the proposed adapted bidirectional Kalman filter method. (c), (f) and (i) are obtained using unadapted bidirectional Kalman filter method.

The results show that in multi-frequency operation, the adapted bidirectional Kalman filter method has the lowest RMSE. The proposed method also performs better than the DFT method in 70% of instances, while the unadapted bidirectional Kalman filter method performs better in only 14% of t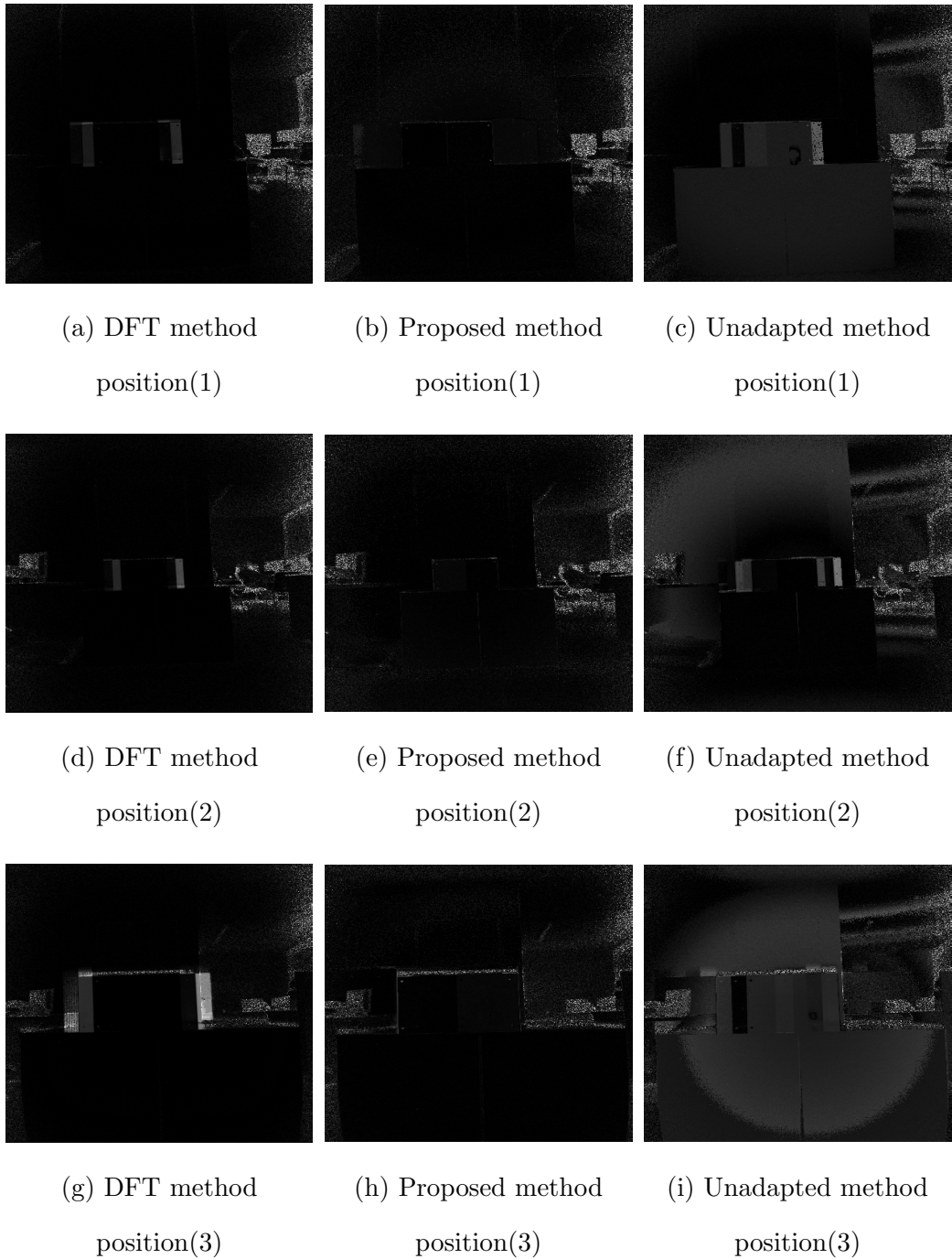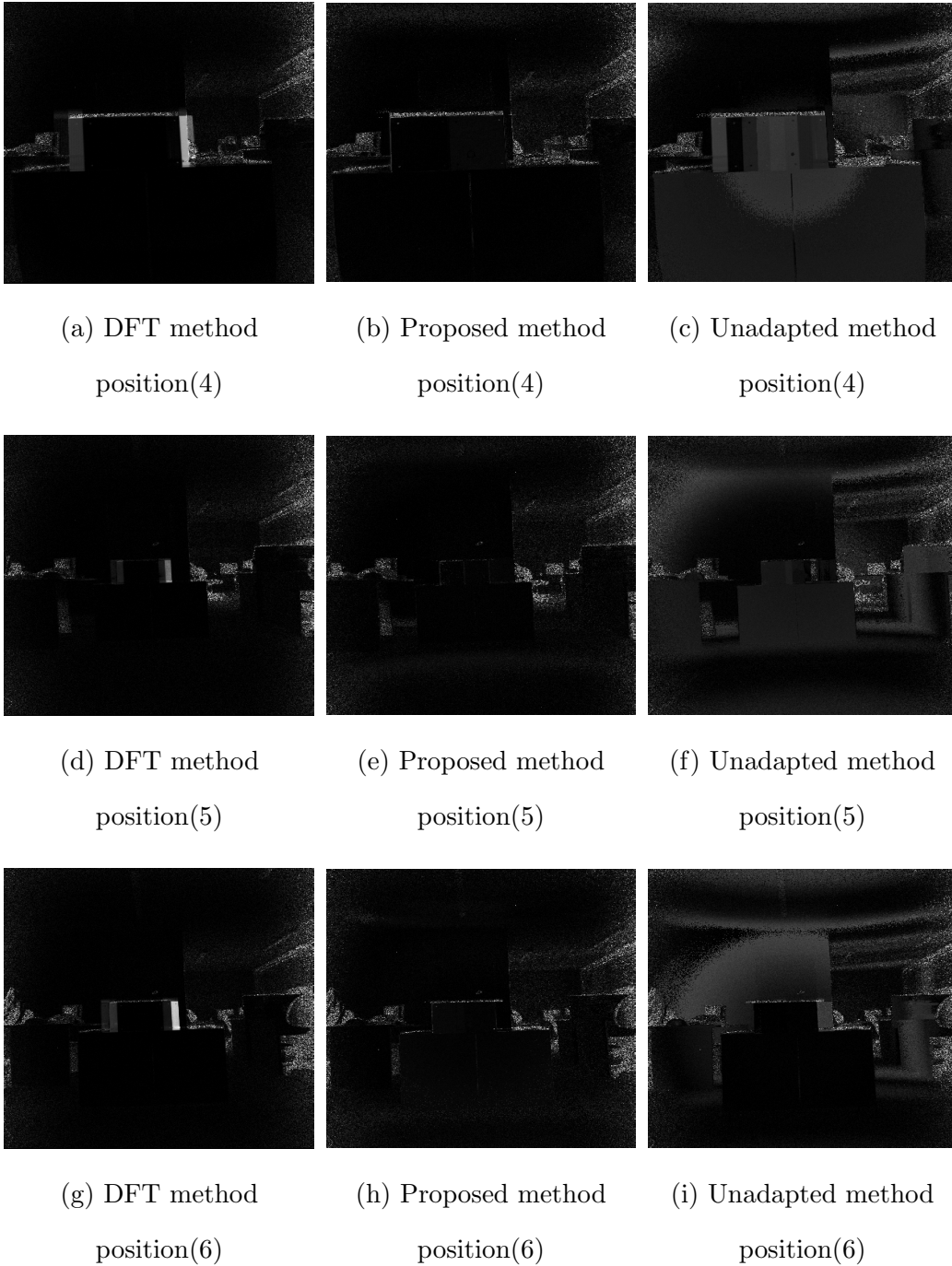he instances. The adapted bidirectional Kalman filter running on multi-frequency raw frames is not as performant as the unadapted bidirectional Kalman filter running on single modulation frequency raw frames. The increased error in multi-frequency operation is expected due to nonlinear effects such as aliasing and multi-path interference

Table 4.2: Comparison of methods. RMSE in units of radians. % is the proportion of pixels with less error than the DFT method.

Method	%	RMSE
Unadapted Bidirectional Kalman Filter (Multi-frequency)	13.57%	0.789
Adapted Bidirectional Kalman Filter (Multi-frequency)	70.35%	0.155
Unadapted Bidirectional Kalman Filter (16 MHz)	77.1%	0.083
Adapted Bidirectional Kalman Filter (16 MHz)	77.06%	0.085
DFT	N/A	0.627

The qualitative experiment contains inter-frame motion, as it is not made from the an artificial set of frames, as such, it is more indicative of the performance of the proposed method in a real-time application. When comparing the phase images produced from the qualitative experiment in Fig. 4.10, we observe that the phase image produced by the adapted bidirectional Kalman filter has the least motion blur of the tested methods.

(a) Traditional DFT processing
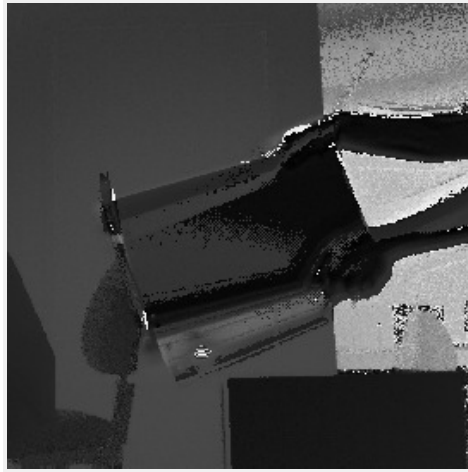
(b) Unadapted bidirectional Kalman filter processing

(c) Proposed bidirectional Kalman filter processing

Figure 4.10: Qualitative comparison of phase images recovered from multi-frequency time-of-flight data taken of a moving object. The proposed method mitigates the motion artefacts visible in the earlier methods.

Chapter 5

Conclusion

Time-of-flight range imaging is especially affected by motion due to its need
for multiple raw frame measurements to construct a single range image. Dy-
namic scenes cause inconsistencies in the raw frame measurements and result
in erroneous range estimates. No method currently exists to reduce motion
error in multi-frequency operated Time-of-light cameras. The time-of-flight
bidirectional Kalman Filter is a state of the art method that reduces error
in range measurements due to transverse motion. The task which this thesis
attempted to achieve is the adaptation of the bidirectional Kalman filter for
use in time-of-flight cameras operating with multiple modulation frequencies.
A key component of the bidirectional Kalman filter is the prediction of the
signal between iterations. In single modulation frequency operation, this is
achieved by making a static scene assumption, which allows the prediction to
be equal to the previous estimation. However, predicting the same signal be-
tween raw frames in multi-modulation frequency operation is not valid even if
the assumption of a static scene made. The prediction is not valid because the
phase shift and the amplitude of the signal vary significantly between mod-
ulation frequencies. A successful prediction of the signal across modulation
frequencies is key to a successful adaptation of the bidirectional Kalman filter.
The rise time of pixel modulation is nonzero. As such, increasing modulation
frequency reduces the amplitude of the correlation waveform. The change in

amplitude due to a change in modulation frequency is modelled linearly, and the linear correlation coefficient is calculated and used to predict the amplitudes across frequencies. The change in phase shift between two frequencies is dependent on the ratio of the two frequencies; If the phase shift measured using the first modulation frequency is beyond 2π, the phase shift appears smaller due to phase wrapping. Operating the time-of-flight Kalman filter across multiple frequencies on phase wrapped data, therefore, incurs nonlinear error. Shifting the phase by an integer multiple of 2π removes this error, but creates a new ambiguity as the number of multiples of 2π that must be accounted for is a priori-unknown. The ambiguity due to phase wrapping is solved by selecting the number of phase wraps that best predicts the next two raw frames. An unintended consequence of the unwrapping method is that it guarantees a low raw frame prediction even when the final predicted phase shift is incorrect. The composite phase shift selection method is changed such that it also bases its decision on raw frame predictions that are not minimised by phase unwrapping methodology. When a qualitative experiment is performed by swinging a board in front of a multi-frequency time-of-flight camera, the traditional DFT method shows significant blurring at the sides of the board where transverse motion is expected. When the bidirectional Kalman filter is used without adapting it to the multi-frequency operation, the blurring becomes worse. However, when the adapted bidirectional Kalman filter is used, the blur due to motion is reduced significantly. When a quantitative experiment is performed by capturing raw frames of a board at multiple locations and simulating motion, the unadapted bidirectional Kalman filter performed better than the DFT method in only 13.6% of the motion-affected pixels, while the proposed bidirectional Kalman filter performed better than the DFT in 70% of the motion affected pixels.

5.1 Limitations and Future Work

Although the adapted bidirectional Kalman filter reduces error significantly in dynamic scenes compared to the traditionally used DFT method, more accurate phase estimations are calculated using the DFT in non-dynamic scenes. The reduction in accuracy is believed to be caused by non-linear effects such as multi-path interference between modulation frequencies which are not accounted for by the adapted bidirectional Kalman filter. Future work could reduce such error by expanding the phase shift prediction equation to take multi-path into account. An alternative to reducing the error in pixels not affected by multi-path is to detect the pixels affected by transverse motion in an image and only apply the adapted bidirectional Kalman filter method to motion affected pixels, while pixels unaffected by transverse motion are processed using the DFT method. Future work could also include an investigation of what modulation frequencies are optimal for more accurate phase estimation using the adapted bidirectional Kalman filter. Such recommendations could be taken into account when manufacturing time-of-flight cameras operating across multiple modulation frequencies.

References

Bamji, C. S., P. O'Connor, T. Elkhatib, S. Mehta, B. Thompson, L. A. Prather, D. Snow, O. C. Akkaya, A. Daniel, A. D. Payne, T. Perry, M. Fenton, and V. Chan. A 0.13 μm CMOS system-on-chip for a 512×424 time-of-flight image sensor with multi-frequency photo-demodulation up to 130 MHz and 2 GS/s ADC. *IEEE Journal of Solid-State Circuits*, **50(1)**, pp. 303–319 (2015).

Blake, J., C. Kerl, F. Echtler, and L. Xiang. libfreenect2: open-source library for Kinect v2 depth camera, release 0.1.1 (2016). `https://zenodo.org/record/50641`.

Dorrington, A. A., M. J. Cree, D. A. Carnegie, A. D. Payne, R. M. Conroy, J. P. Godbaz, and A. P. Jongenelen. Video-rate or high-precision: A flexible range imaging camera. In: *Image Processing: Machine Vision Applications*, volume 6813, p. 681307. International Society for Optics and Photonics (2008).

Foix, S., G. Alenya, and C. Torras. Lock-in time-of-flight (ToF) cameras: A survey. *IEEE Sensors Journal*, **11(9)**, pp. 1917–1926 (2011).

Fuchs, S. Multipath interference compensation in time-of-flight camera images. In: *2010 20th International Conference on Pattern Recognition*, pp. 3583–3586 (2010).

Godbaz, J. P., M. J. Cree, and A. A. Dorrington. Closed-form inverses for the mixed pixel/multipath interference problem in AMCW lidar. In: *Computational Imaging X*, edited by C. A. Bouman, I. Pollak, and P. J. Wolfe,

volume 8296, pp. 302 – 316. International Society for Optics and Photonics, SPIE (2012).

Goh, Y. H., P. Raveendran, and Y. L. Goh. Robust speech recognition system using bidirectional Kalman filter. *IET Signal Processing*, **9(6)**, pp. 491–497 (2015).

Grzegorzek, M., C. Theobalt, R. Koch, and A. Kolb. *Time-of-flight and depth imaging. Sensors, algorithms and applications: Dagstuhl Seminar 2012 and GCPR Workshop on Imaging New Modalities*, volume 8200. Springer (2013).

He, Y. and S. Chen. Recent advances in 3D data acquisition and processing by time-of-flight camera. *IEEE Access*, **7**, pp. 12495–12510 (2019).

Hu, X., D. V. Prokhorov, and D. C. W. II. Time series prediction with a weighted bidirectional multi-stream extended Kalman filter. *Neurocomputing*, **70(13-15)**, pp. 2392 – 2399 (2007).

Jimenez, D., D. Pizarro, and M. Mazo. Single frame correction of motion artifacts in PMD-based time of flight cameras. *Image and Vision Computing*, **32(12)**, pp. 1127 – 1143 (2014).

Kalman, R. A new approach to linear filtering and prediction problems. *Transactions of the ASME - Journal of Basic Engineering*, **82**, pp. 35–45 (1960).

Karel, W., S. Ghuffar, and N. Pfeifer. Modelling and compensating internal light scattering in time of flight range cameras. *The Photogrammetric Record*, **27(138)**, pp. 155–174 (2012).

Lange, R. *3D time-of-flight distance measurement with custom solid-state image sensors in CMOS/CCD-technology*. Ph.D. thesis, Universität Siegen (2000).

Lange, R. and P. Seitz. Solid-state time-of-flight range camera. *IEEE Journal of Quantum Electronics*, **37(3)**, pp. 390–397 (2001).

Lefloch, D., T. Hoegg, and A. Kolb. Real-time motion artifacts compensation of ToF sensors data on GPU. In: *Three-Dimensional Imaging, Visualization, and Display 2013*, edited by B. Javidi and J.-Y. Son, volume 8738, pp. 166 – 172. International Society for Optics and Photonics, SPIE (2013).

Lindner, M. and A. Kolb. Compensation of motion artifacts for time-of-flight cameras. In: *Workshop on Dynamic 3D Imaging*, pp. 16–27. Springer (2009).

Lottner, O., A. Sluiter, K. Hartmann, and W. Weihs. Movement artefacts in range images of time-of-flight cameras. In: *2007 International Symposium on Signals, Circuits and Systems*, volume 1, pp. 1–4. IEEE (2007).

Payne, A. D., A. A. Dorrington, M. J. Cree, and D. A. Carnegie. Improved linearity using harmonic error rejection in a full-field range imaging system. In: *Three-Dimensional Image Capture and Applications 2008*, edited by B. D. Corner, M. Mochimaru, and R. Sitnik, volume 6805, pp. 101 – 111. International Society for Optics and Photonics, SPIE (2008).

Streeter, L. Methods for linear radial motion estimation in time-of-flight range imaging. In: *Videometrics, Range Imaging, and Applications XIV*, volume 10332, p. 103320C. International Society for Optics and Photonics (2017a).

Streeter, L. Stochastic calculus analysis of optical time-of-flight range imaging and estimation of radial motion. *JOSA A*, **34(7)**, pp. 1063–1072 (2017b).

Streeter, L. Time-of-flight range image measurement in the presence of transverse motion using the Kalman filter. *IEEE Transactions on Instrumentation and Measurement*, **67(7)**, pp. 1573–1578 (2018).

Streeter, L., M. J. Cree, and A. A. Dorrington. A strategy for the correction of effects of jitter in AMCW lidar images. In: *2013 28th International Conference on Image and Vision Computing New Zealand (IVCNZ 2013)*, pp. 500–505. IEEE (2013).

Streeter, L. and A. A. Dorrington. Coded exposure correction of transverse motion in full-field range imaging. *Optical Engineering*, **53**, p. 102109 (2014).

Streeter, L. and A. A. Dorrington. Simple harmonic error cancellation in time of flight range imaging. *Optics letters*, **40(22)**, pp. 5391–5394 (2015).

Streeter, L. and Y. C. Kuang. Metrological aspects of time-of-flight range imaging. *IEEE Instrumentation Measurement Magazine*, **22(2)**, pp. 21–26 (2019).