Real-time Image Stabilization for ToF Cameras on Mobile Platforms

Benjamin Langmann, Klaus Hartmann, and Otmar Loffeld

Center for Sensor Systems (ZESS), University of Siegen, Paul-Bonatz-Str. 9-11, 57068 Siegen, Germany, {langmann, hartmann, loffeld}@zess.uni-siegen.de

Abstract. In recent years, depth cameras gained increasing acceptance in the areas of robotics and autonomous systems. However, on mobile platforms depth measurements with continuous wave amplitude modulation Time-of-Flight cameras suffer from motion artifacts, since multiple acquisitions are required in order to compute one depth map (resulting in longer effective exposure times). Some lenses of different manufacturers include image stabilizers, but they are only able to compensate for small image shifts. Moreover, when performing a phase unwrapping based on the acquisition of multiple depth maps with different modulation frequencies, the motion artifacts are significantly more severe. In this paper, a method to compensate camera motions during the acquisition of a single depth map as well as for multiple depth maps is presented. Image shifts are estimated firstly and after normalization the individual phase images are shifted accordingly. The proposed approach is evaluated on different scenes and it is able to facilitate ToF imaging on mobile platforms.

Keywords: Motion Compensation, Time-of-Flight, 3D Imaging, Photonic Mixer Device (PMD)

1 Introduction

Depth imaging devices act as an important sensor in the areas of robotics, autonomous systems and similar research fields. However, as camera motions during an image acquisition result in a blur for color imaging, they result in motion artifacts for depth imaging. These artifacts consist of invalid depth measurements and can in general assume any value. Most Time-of-Flight depth cameras, e.g. based on the Photonic Mixer Device (PMD), utilize multiple images namely phase images to compute one depth map. Motion artifacts are here caused by camera motions during the acquisition of a single phase image and in between the acquisitions of the phase images. The former case can usually be neglected due to the relatively short exposure times which are commonly applied. Since typical acquisition times for all phase images required to obtain one depth map lie between 10 ms and 50 ms, the camera can be subject to significant motion during that time on mobile platforms. Two features in the scene induce motion artifacts for continuous wave amplitude modulation ToF depth cameras. The first one are of course edges, which result in measuring different distances, i.e. phase shifts. Secondly, different reflectivities introduce motion artifacts additionally and the reason for this will be discussed later on.

An approach to compensate camera motions in between multiple phase images is proposed in this paper. It works by firstly estimating the lateral camera motion in between phase images and then by shifting the phase images accordingly. Several peculiarities of PMD chips have to be considered and will be discussed in the course of this paper. Moreover, when a multi-frequency phase unwrapping approach is applied in order to extend the measurement range, motion artifacts become a severe limitation. Therefore, we extend the proposed image stabilization technique to multiple depth maps.

The paper is structured as follows. The related research is discussed in Section 2 and in Section 3 the hardware used in this paper is presented. Afterwards, the proposed method is presented in Section 4. Experiments to demonstrate the capabilities and limitations of the proposed approach are reviewed in Section 5 and the paper ends with a conclusion in Section 6.

2 Related Work

In [1] Lottner et al. analyze motion artifacts encountered for sufficiently fast motion of objects in the scene when using PMD cameras. Lindner and Kolb introduce in [2] motion compensation based on optical flow for moving objects in videos acquired with PMD cameras. In an industry environment motion compensation is demonstrated by Hussmann et al. in [3] and [4]. They perform the motion compensation on PMD depth maps and the motion is detected by comparison of binary foreground maps. Axial motion of ToF cameras is discussed in [5] using the Windowed Discrete Fourier Transform.

Unlike most previous approaches, the proposed method does not treat motion artifacts like so-called flying pixels and tries to remove them by estimating local motion. These approaches produce visibly nice results but may also result in incorrect contours of objects (since a flying pixel is caused by a mixture of multiple distances). Instead, whole phase images are shifted in relation to each other in order to account for the observed camera motion. This method is computationally inexpensive and does not introduce incorrect measurements. The approach neglects camera rotations around the roll axis, since this kind of rotation typically causes small effects for short acquisition times and since rotating a phase image to compensate it would introduce aliasing.

Phase unwrapping to remove ambiguities in depth measurements is a common task in several research areas, e.g. remote sensing. In the context of depth imaging, a probabilistic approach to remove ambiguities in a single depth image is proposed in [6]. An optimization is performed based on a cost function aiming at the removal of discontinuities. In [7] this method is extended to incorporate multiple measurements with different modulation frequencies and hence different ambiguity ranges. A different approach based on a single depth map is introduced in [8]. The unambiguous depth of an object is inferred here by observing how much infrared light it reflects. The approach does not handle each pixel individually, but finds edges in the depth map in order to account for the different reflectivity of objects, which would otherwise compromise the results. Phase unwrapping methods are in general not completely stable and a method to obtain smooth results in subsequent phase unwrapped depth maps is described in [9]. Moreover, in [10] phase unwrapping is performed with the help of stereo information.

3 ZESS MultiCam with Medium-Range Lighting

Recently, a new version of the MultiCam (cf. [11]) equipped with Gigabit Ethernet, a 3 megapixel color CMOS chip and the 19k PMD chip of PMD Technologies was developed. Both chips share the same lens utilizing a Bauernfeind prism with an integrated beam splitter. The heart of the camera is a Xilinx FPGA chip and the camera features a C-mount lens adapter. See Fig. 1 for a picture and specifications of the MultiCam.

We developed a medium-range lighting system, which can be easily scaled up,



Fig. 1. The MultiCam, a 2D/3D monocular camera, and its specifications.

and it was designed to capture large viewing angles, e.g. 40 degrees. The light is produced by chip LEDs with a maximum continuous optical power of 3.5 Watt, which are available from Osram (SFH-4750). These LEDs have an emission peak at 860 nm, the active area is relatively dense unlike large LED arrays and they feature low and symmetric rise and fall times of only 10 ns. Three prototypes with up to 26 LEDs with 3.5 Watt continuous optical power each are shown in Fig. 2. A collimator with a half angle of 11 degrees is mounted on each LED and the LEDs are adjusted to cover the observed area. With this lighting systems we can capture scenes up to 70 meters for smaller viewing angles and depending on the exposure times possible. The device in Fig. 2c is mounted on a rotary table, which was used to simulate camera motions in the experiments.

Measurement results of an outdoor scene are shown in Fig. 3 using only 8 LEDs.

The distance to the building is 50 meters. The results demonstrate the capabilities of this approach with weaknesses at small objects and highly structured objects due to the limited lateral resolution of the PMD chip.



(a) Compact 10 LED de- (b) Flexible 20 LED de- (c) vice vice tar

(c) 26 LED device on a rotary table

Fig. 2. Three medium-range development prototypes with 10, 20 and 26 LEDs.

4 ToF Image Stabilization

Continuous wave amplitude modulation based ToF imaging operates normally by acquiring four images under different phase shift. Motion artifacts in the computed depth map occur obviously when measurement samples of objects with different distances are combined resulting in a so-called mixed phase. However, when working with PMD chips, motion artifacts additionally occur for different intensities, e.g. on a checkerboard. Different effective offsets and gains of both channels of PMD pixels caused by significant variations in the chip will not get canceled out in the four phase algorithm as when all samples originate from one PMD pixel. For details on PMD imaging see [12] or [13].

The proposed acquisition procedure is outlined as follows: Firstly, the phase images are normalized using the calibration method described in Section 4.1. Then the lateral image shifts between the first phase image and all three consecutive phase images, which are caused by rotations around and motions on the pitch and jaw axis of the camera (perspective effects can be neglected), are estimated with the method detailed in Section 4.2. Afterwards, the phase images are shifted and the standard PMD computations to obtain depth, modulation amplitude and grayscale images are performed as described in Section 4.3. Ambiguities of depth measurements in larger scenes can be resolved with the phase unwrapping approach discussed in Section 4.4 while compensating for these camera motions.



Fig. 3. Measurement results obtained with the medium-range lighting using 8 LEDs. A modulation frequency of 20 MHz and an exposure time of 6 ms were used and the distance to the building is 50 meters.

4.1 Intensity Calibration for PMD Chips

When camera motions are compensated by combining measurements of different pixels, e.g. by using shifted phase images, pixel values must be comparable. However, this requires an intensity calibration of all PMD pixels to account for the fixed pattern noise. The same calibration is necessary in most applications, in which the grayscale image of a PMD camera is utilized in a given task. In Fig. 4 the raw intensity values for one channel of a phase image are shown. Current PMD imaging chips utilize three AD converters with different effective offsets and gains, which results in vertical lines in addition to pixel variations. In order to perform the normalization, a set of images with different exposure times of a uniformly lit diffuse glass is acquired. The relative offset and gain for both channels of each pixel can be estimated based on these images, which enables an affine normalization of the phase images. The resulting intensity values for one channel are also given in Fig. 4.

The exact calibration procedure is detailed in the following. Let Γ_i for $\Gamma \in [A, B]$ and i = 1, ..., n be n phase images for a PMD channel Γ . The index i describes here different acquisitions, since the phase is not of importance now. The pixels



Fig. 4. Affine normalization of a PMD phase image. Left: original phase image for the first PMD channel, right: normalized image.

of the image are denoted by $\Gamma_i^{(x,y)}$ and let t_i be the associated integration time. Then the average pixel value $\mu_{\Gamma}^{(x,y)}$ is given by

$$\mu_{\Gamma}^{(x,y)} = \frac{1}{n} \sum_{i=1}^{n} \Gamma_i^{(x,y)} .$$
 (1)

The affine parameters for each channel $\alpha_{\Gamma}^{(x,y)}$ and $\beta_{\Gamma}^{(x,y)}$ are calculated with

$$\tau_i = t_i - \frac{1}{n} \sum_{j=1}^n t_j \tag{2}$$

$$\alpha_{\Gamma}^{(x,y)} = \frac{\sum_{i=1}^{n} \tau_i \cdot \left(\Gamma_i^{(x,y)} - \mu_{\Gamma}^{(x,y)}\right)}{\sum_{i=1}^{n} \tau_i^2}$$
(3)

$$\beta_{\Gamma}^{(x,y)} = \mu_{\Gamma}^{(x,y)} - \alpha_{\Gamma}^{(x,y)} \frac{1}{n} \left(\sum_{j=1}^{n} t_j \right) .$$
 (4)

Now the normalized pixel value $\hat{\Gamma}^{(x,y)}$ of a PMD channel Γ for an acquired pixel value $\Gamma^{(x,y)}$ can be computed with

$$\mu_{\Gamma}^{\alpha} = \frac{1}{|\Gamma|} \sum_{(x,y)\in\Gamma} \alpha_{\Gamma}^{(x,y)} \tag{5}$$

$$\hat{\Gamma}^{(x,y)} = \left(\Gamma^{(x,y)} - \beta_{\Gamma}^{(x,y)}\right) \cdot \frac{\mu_{\Gamma}^{\alpha}}{\alpha_{\Gamma}^{(x,y)}} .$$
(6)

All recent PMD chips feature a mechanism to remove ambient light (Suppression of Backlight Illumination - SBI), which enables outdoor imaging. However, the SBI obviously also effects the intensity values, which will disturb motion estimates based on these intensity images. In order to characterize this behavior, a dataset consisting of 100 images for integration times between 0.1 ms and 5 ms with a step size of 0.1 ms was acquired. The raw values for both channels A and B as well as for the four phase images are displayed for one pixel in Fig. 5a. One can clearly observe the linear region of the PMD chip up to 0.8 ms. The SBI is active for higher integration times. The PMD chip enters the saturation region for one phase image when applying integration times larger than 2.5 ms. In 5b the resulting phase intensity values $I_i^{(x,y)} = A_i^{(x,y)} + B_i^{(x,y)}$ are plotted. In order to obtain increasing intensity values a correction was applied as follows. If the value $\Gamma_i^{(x,y)}$ of one channel $\Gamma \in [A,B]$ of a PMD pixel exceeds a threshold, e.g. $\gamma_{thres} = 36500$, a normalization term of $3 \cdot |A_i^{(x,y)} - B_i^{(x,y)}|$ is added to the phase intensity $I_i^{(x,y)}$. If the intensity response of a PMD chip is known for a given lens, the real intensity can be determined for an exposure time and modulation amplitude.



Fig. 5. Response curves of a PMD chip and the calculated as well as corrected intensity for different exposure times. Shown are the raw values for both channels of a single pixel and four phase offsets in the first graph and the intensity values using the standard formula as well as with SBI correction in the second.

4.2 Motion Estimation

Vandewalle et al. introduced a method for planar motion estimation based on the Fourier Transformation in [14]. Since fast rotations around the roll axis are rarely observed in typical situations, the method is reduced to lateral image shifts for an increased robustness as well as to avoid aliasing when rotating images. Let $f(\underline{x})$ be the reference image and $g(\underline{x})$ the actually observed one with $g(\underline{x}) =$ $f(\underline{x}+\Delta \underline{x})$ with Fourier Transformations $F(\xi)$ and $G(\xi)$. Then the following holds

$$G(\xi) = \int g(\underline{x}) e^{-i2\pi\xi^T \cdot \underline{x}} d\underline{x}$$
(7)

$$= \int f(\underline{x} + \Delta \underline{x}) e^{-i2\pi\xi^T \cdot \underline{x}} d\underline{x}$$
(8)

$$= \int f(\underline{\hat{x}}) e^{-i2\pi\xi^T \cdot (\underline{\hat{x}} - \Delta \underline{x})} d\underline{\hat{x}}$$
(9)

$$=e^{i2\pi\xi^T\cdot\underline{\Delta}\underline{x}}\cdot\int f(\underline{\hat{x}})e^{-i2\pi\xi^T\cdot\underline{\hat{x}}}d\underline{\hat{x}}$$
(10)

$$=e^{i2\pi\xi^T\cdot\underline{\Delta}\underline{x}}\cdot F(\xi) \tag{11}$$

with a substitution $\underline{\hat{x}} = \underline{x} + \Delta \underline{x}$.

Therefore, the phase difference $e^{i2\pi\xi^T\cdot\Delta\underline{x}}$ can be obtained given $F(\xi)$ and $G(\xi)$ with

$$\xi^T \cdot \underline{\Delta x} = \frac{1}{2\pi} \arg\left(\frac{G(\xi)}{F(\xi)}\right) . \tag{12}$$

Instead of computing $\Delta \underline{x}$ with for a single ξ , Vandewalle et al. argue to construct a system of linear equations for a set of frequencies and to use the least squares estimate for $\Delta \underline{x}$ in order to avoid aliasing.

4.3 Intra-Frame Motion Compensation

High accuracy motion compensation for dynamic scenes including local and global motion is difficult to perform in real-time, in particular when it serves as a pre-processing step and other image processing tasks are should be processed in parallel. In order to accomplish motion compensation (MC) in real-time, we simplify the task by considering only global motions, neglecting roll and by restricting image shifts to full pixels. Let A_i and B_i with i = 1, 2, 3, 4 be the phase images for PMD channel A and B respectively acquired with the *i*-th phase shift and let $I_i = A_i + B_i$ be the phase intensity image. With the motion estimation method detailed in Section 4.2 the lateral image shifts Δ_i between I_1 and I_i for i = 2, 3, 4 can be determined and rounded to the nearest integer. These image shift are applied to the associated phase images (can be replaced with pointer arithmeticians) and the PMD calculations for demodulation are performed.

4.4 Phase Unwrapping

Continuous wave amplitude modulation Time-of-Flight depth measurements do not measure distances directly, but derive the distance from the phase difference between emitted and received light. This means that the distance measured is ambiguous and multiples of the half wavelength $\lambda = \frac{c}{2\nu}$ with c being the speed of light can be added. If the observed scene contains longer distances than λ and if the modulation frequency should not be decreased in order the to maintain the depth resolution, a method to perform a phase unwrapping is required to retrieve correct distances. In the following, a method which combines the depth maps obtained with two different modulation frequencies is presented, which extends the unambiguous range to the lowest common multiple of both half wavelengths. Let d_1, d_2, \ldots, d_n be distances measured with modulation frequencies $\nu_1, \nu_2, \ldots, \nu_n$. Then the most likely absolute distance $\hat{d}_i = d_i + k_i \frac{c}{2\nu_i}$ of all measurements combined can be obtained by minimizing the squared difference of all proposed real distances

$$min_{k_1,\dots,k_n \in \mathcal{N}_0} \left\{ \sum_{i=1}^{n-1} \sum_{j=i+1}^n \left(d_i + k_i \frac{c}{2\nu_i} - d_j - k_j \frac{c}{2\nu_j} \right)^2 \right\} .$$
(13)

The obvious approach is to combine a small frequency (e.g. 1 MHz with an ambiguity range of 150 m) and a larger frequency with a high depth resolution. However, this poses a significant challenge to the lighting devices. In practice, the frequencies must be similar to avoid damages and unstable behavior or low power lighting. If the size of the set is small, e.g. just two, the minimization can be performed by just calculating all possibilities (up to a given maximum distance).

Motion artifacts are even more prominent when phase unwrapping is conducted. However, the motion compensated depth maps acquired with two (or more) modulation frequencies can also be motion compensated (inter-frame MC). The camera motion can then be estimated based on both first phase intensity images I_1 and the depth maps are shifted accordingly. Alternatively, the camera motion can be estimated based on all four pairs of phase intensity images by averaging the image shifts.

5 Experiments

In Fig. 6 an example for the intra-frame MC is given for a test setup acquired with an exposure time of $5 \,\mathrm{ms}$, a $50 \,\mathrm{mm}$ lens and two different modulation fre-



Fig. 6. Results for real-time motion compensation experiment involving a moving camera and two modulation frequencies.



Fig. 7. Demonstration of severe motion artifacts when performing phase unwrapping. Results for normal and motion compensated (MC) phase unwrapping based on standard and motion compensated depth maps.

quencies. A motion between phase images of about 1.2 pixels is estimated and subsequent images with different modulation frequencies are shown. The results demonstrate that the motion artifacts can be largely reduced. However, at the edges of objects and high contrast regions some artifacts persist, since only full pixel shift are applied.

In Fig. 7 four cases to perform the phase unwrapping are considered: Without motion compensation, with two motion compensated depth maps, with two normal depth maps but with inter-frame MC and finally with motion compensated depth maps and inter-frame MC. The results demonstrate how important the proposed method is for medium-range depth imaging with phase unwrapping on mobile platforms or when other camera motions occur.

A more realistic experiment to demonstrate the intra-frame MC is displayed in Fig. 8. Here a modulation frequency of 20 MHz, a 8.5 mm lens and an exposure time of 10 ms were used. A fast motion was introduced and results in motion artifacts at the edges of objects, which are successfully removed when applying the MC. Only flying pixels at the edges of objects remain, which cannot be corrected without affecting the contours of objects. However, a simple thresholding with the modulation amplitude usually allow to remove them.



Fig. 8. Experiment to evaluate the intra-frame MC using a single modulation frequency of 20 MHz.

Results for another experiment to demonstrate the inter-frame MC with a different camera, lens and illumination system are depicted in Fig. 9. The camera and the illumination system were mounted on a rotation table and depth maps were acquired with alternating modulation frequencies of 17 and 19 MHz and an exposure time of 5 ms. The results show how the incorrect estimation of the real distances can be reduced at discontinuities, e.g. at the doors. It should be mentioned here that we observe multi-path reflections on the reflective floor.

The inter-frame is not effective for slow camera motions and short acquisition times. The angular velocity v which results in an image shift of one pixel between the first and the last phase image can serve as a criterion and it is calculated with

$$v = \frac{\alpha_{\rm cam}}{N \cdot t_{\rm acq}} \tag{14}$$

in degrees per second for an opening angle of $\alpha_{\rm cam}$, N pixels in direction of the rotation and an acquisition time of $t_{\rm acq} = 3 * (t_{\rm exp} + t_{\rm ro})$. Here $t_{\rm exp}$ is the exposure time for a phase image and $t_{\rm ro}$ is the read-out time of the PMD chip. In the experiments we get for a horizontal rotation (yaw axis), the 19k PMD chip and an exposure time of 5 ms a velocity of $14.2^{\circ}/s$ for a 8.5 mm lens and for a 50 mm lens a critical angular velocity of $2.6^{\circ}/s$.



Fig. 9. Phase unwrapping experiment on a rotation table. An image shift of 3 pixels was detected between the first phase images of both acquisitions.

6 Conclusion

Operating a depth camera on a mobile platform poses challenges in general and for ToF depth cameras motions of the camera during the acquisition on a depth map results in motion artifacts. The longer the acquisition time is, the more severe are the motion artifacts. This is in particular a problem when measuring larger scenes due to required larger exposure times or even when fusing multiple depth maps for phase unwrapping as well as in order to reduce noise. In this paper, real-time methods to reduce these motion artifacts significantly are presented. Based on methods to calibrate the fixed pattern noise and to estimate image shifts, phase images are aligned before processing in order to reduce the motion artifacts originating from global motions in the images. Similarly, multiple depth maps are fused with the proposed approach to obtain depth measurements of larger scenes. The approach was evaluated with several experiments to confirm its capabilities and in order to decide if the methods are required for a given application.

Acknowledgments

This work was funded by the German Research Foundation (DFG) as part of the research training group 1564 "Imaging New Modalities".

References

- Lottner, O., Sluiter, A., Hartmann, K., Weihs, W.: Movement artefacts in range images of time-of-flight cameras. In: Signals, Circuits and Systems, 2007. ISSCS 2007. International Symposium on. Volume 1. (2007) 1–4
- Lindner, M., Kolb, A.: Compensation of motion artifacts for time-of-flight cameras. In: Proc. Dynamic 3D Imaging. LNCS, Springer (2009) 16–27
- Hussmann, S., Hermanski, A., Edeler, T.: Real-time motion supression in tof range images. In: Proc. IEEE Instrumentation and Measurement Technology Conf. (I2MTC). (2010) 697–701
- Hussmann, S., Hermanski, A., Edeler, T.: Real-time motion artifact suppression in tof camera systems. IEEE Trans. Instrum. Meas. 60(5) (2011) 1682–1690
- Drayton, B., Carnegie, D., Dorrington, A.: Phase algorithms for reducing axial motion and linearity error in indirect time of flight cameras. Sensors Journal, IEEE **PP**(99) (2013) 1–1
- Droeschel, D., Holz, D., Behnke, S.: Probabilistic Phase Unwrapping for Time-of-Flight Cameras. In: Proceedings of the joint conference of the 41st International Symposium on Robotics (ISR 2010) and the 6th German Conference on Robotics (ROBOTIK 2010), Munich, Germany (June 2010) 318–324
- Droeschel, D., Holz, D., Behnke, S.: Multi-frequency phase unwrapping for timeof-flight cameras. In: Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on. (2010) 1463–1469
- McClure, S.H., Cree, M.J., Dorrington, A.A., Payne, A.D.: Resolving depthmeasurement ambiguity with commercially available range imaging cameras. Proc. SPIE 7538, Image Processing: Machine Vision Applications III (2010) 75380K– 75380K–12
- 9. Choi, O., Lee, S., Lim, H.: Interframe consistent multifrequency phase unwrapping for time-of-flight cameras. Optical Engineering **52**(5) (2013) 057005–057005
- Choi, O., Lee, S.: Fusion of time-of-flight and stereo for disambiguation of depth measurements. In: Proceedings of the 11th Asian conference on Computer Vision - Volume Part IV. ACCV'12, Berlin, Heidelberg, Springer-Verlag (2013) 640–653
- Prasad, T., Hartmann, K., Weihs, W., Ghobadi, S., Sluiter, A.: First steps in enhancing 3d vision technique using 2d/3d sensors. In: Computer Vision Winter Workshop, Telc, Czech Republic (2006) 82–86
- 12. Möller, T., Kraft, H., Frey, J., Albrecht, M., Lange, R.: Robust 3d measurement with pmd sensors. Range Imaging Day, Zürich Section 5 (2005)
- Lottner, O., Weihs, W., Hartmann, K., for Sensor, C., Systems: Systematic nonlinearity for multiple distributed illumination units for time-of-flight (pmd) cameras. In: 12th WSEAS International Conference on SYSTEMS, Heraklion, Greece, July 22-24, 2008. (2008)
- Vandewalle, P., Suesstrunk, S., Vetterli, M.: A Frequency Domain Approach to Registration of Aliased Images with Application to Super-Resolution. EURASIP Journal on Applied Signal Processing (special issue on Super-resolution) 2006 (2006)