MULTI-CAMERA INTERFERENCE CANCELLATION OF TIME-OF-FLIGHT (TOF) CAMERAS

Lianhua Li[1,2], *Sen Xiang*[1], *You Yang*[1], *Li Yu*[1]

[1] School. of Electron. Inf. & Commun., Huazhong Univ. of Sci. & Tech., Wuhan, China
[2] Air Force Early Warning Academy, Wuhan, China
{llhua, xiangsen, yangyou, hustlyu}@hust.edu.cn

ABSTRACT

In the applications based on depth, multiple TOF cameras are often required to capture the same scene. But if multiple cameras operate simultaneously on the same frequency, they interfere with each other. The multi-camera interference causes a lot of errors in depth measurement. The depth quality is severely reduced, which limits the application of TOF cameras and needs to be resolved. In this paper, a multi-camera interference model is presented. The interference signal for multiple frames is proved to be an ergodic and wide-sense stationary stochastic process. The least square estimation of noninterference signal is proposed to remove the multi-camera interference. The results of experiments prove the approach can recover the depth and amplitude information of multiple TOF cameras from the severe interference.

Index Terms— time of light, depth measurement, multi-camera interference

1. INTRODUCTION

With active optical measuring technique, TOF camera can capture the scene from a single viewpoint in real time. It can simultaneously get range and amplitude images at a high frame rate. Due to these desirable features, TOF camera is rapidly applied in numerous fields including navigation, robotics, machine vision and consumer electronics [1][2].

In three dimensional (3D) scenes, the quality is very important for depth information to influence the 3D reconstruction. Most of current TOF cameras emit the near infrared light with amplitude modulated continuous wave (AMCW) to illuminate the scene, and get the depth of the scene point with measuring the phase of the reflection light signal in each pixel. In the application of single camera, the multiple path interference (MPI) causes the depth measurement error in some special cases, which has recently received significant attention [3][4][5][6]. In multiview synthesis applications, multiple cameras are needed to capture the scene. Cameras receive the reflection light signals with the same modulation frequency from other cameras, which is denoted as multi-camera interference. In the case of multiple cameras, the multi-camera interference problem impairs the depth quality severely and is more serious than MPI, which limits the application of the TOF cameras and needs to be solved.

To eliminate the multi-camera interference, the time-division multiplexing can be used. In this case, every TOF camera operates for different time period to avoid interfering with each other. This solution assumes a synchronization among the cameras and needs another synchronous apparatus or makes the cameras work in turn, which is inconvenient. Another possibility is frequency-division multiplexing which make each camera operate at a different modulation frequency. This method can also avoid interference and demands each camera has its own modulation frequency which decides the maximum detection range and measurement accuracy. Due to limited available frequencies, just a few TOF cameras can work simultaneously, making the application non-ideal and inconvenient. The pseudo-noise modulation approaches are proposed in [7][8][9], which make the reflection signals from other cameras uncorrelated, and are available to avoid the multi-camera interference. But these methods need to upgrade most TOF cameras which utilize the sinusoidal modulation technique. Under the condition of sinusoidal modulation, because the interference signals are related to the noninterference signals, it is difficult to recover the truth signal from the multi-camera interference.

In this paper, we propose a novel method to deal with the challenge. By analyzing the statistical property of interference signals for multiple frames, We propose the least squares estimation of noninterference signal (LSENS) to solve the interference problem and recover the depth and amplitude information from the severe interference. Unlike the existing approaches, we use commercial TOF cameras directly without hardware modification.

The remainder of the paper is organized as follows. In Section 2, the principle of TOF camera is explained. Section 3 describes the parametric modeling of multi-camera interference in TOF cameras. Section 4 discusses the proposed ap-

This work is supported in part by the National Natural Science Foundation of China (NSFC) (Grand No. 61231010, 61170194, 61202301), Research Fund for the Doctoral Program (20120142110015).

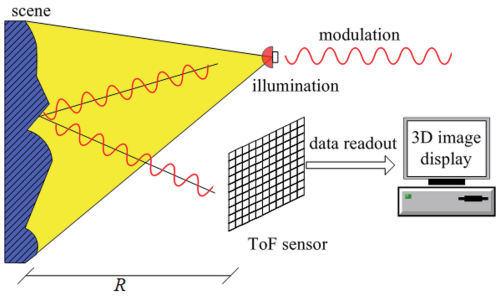


Fig. 1. Principle of TOF Measurement

proach for multi-camera interference cancellation, followed by the experiment results in section 5. Finally, we conclude the paper in Section 6.

2. TOF CAMERA MEASUREMENT MODEL

The measurement principle of TOF camera is shown in Fig.1. Amplitude modulated continuous wave is emitted by an illumination unit and the modulated signal is usually co-located at the sensor as reference signal which can be thought of a sinusoidal signal. It is received by the pixel sensor after reflected by the scene. Both amplitude and range are measured at each pixel, thus capturing the scene's 3D information. In current cameras, using a simple computation [10][11] to express the transmitted signal as

$$s(t) = 1 + cos(2\pi f_0 t) \tag{1}$$

with modulation frequency $f_0 = 1/T_0$, where T_0 is the modulation period. The received signal of the TOF lock-in sensor can be written as

$$r(t) = B + Acos(2\pi f_0 t + \varphi) \tag{2}$$

with phase $\varphi = 4\pi f_0 d/c$, where amplitude A is related to the propagation losses and target reflectivity, B is the offset due to the background or ambient light from the scene and d is the measured range with the light speed c. So the TOF camera can get the amplitude image and depth image of the scene together by estimating the amplitude A and phase φ, which means to demodulate the received signal by a cross correlation with the original modulation signal for some sampling tap t_p, as

$$C(t)\,|_{kt_p} = s(t) \otimes r(t), k \in \mathbb{Z} \tag{3}$$

where $\otimes$ denotes the correlation function.

In practice, the correlation is just a sampling process with equally spaced sampling points of duration Δt. Because most TOF cameras are so called 4-tap devices [6], with 4 correlation measurements per modulation period at sampling steps $t_p = T_0/4, k = 0 \cdots 3$, the camera can compute the approximations $\{A, \varphi\}$ [10], as

$$A = P_0\sqrt{(C(3t_p) - C(t_p))^2 + (C(0) - C(2t_p))^2}/2 \tag{4}$$

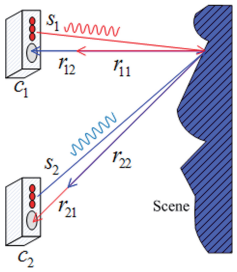


Fig. 2. Multiple TOF Cameras Measurement

$$\varphi = \arctan\left(\frac{C(3t_p) - C(t_p)}{C(0) - C(2t_p)}\right) \tag{5}$$

where P_0 is a constant related with the duration Δt and the modulation frequency f_0. The distance d can be computed as

$$d = c\varphi/4\pi f_0 \tag{6}$$

3. MULTI-CAMERA INTERFERENCE MODEL

When multiple TOF cameras are utilized to capture the scene simultaneously, they are disturbed with each other, if they operate with the same frequency. Without loss of generality, we discuss the case of two TOF cameras. As shown in Fig.2, camera C_1 and C_2 respectively emit sinusoidal modulation light signal s_1 and s_2 with the same modulation frequency onto the scene actively. The pixel sensor of C_1 receives the combined sinusoidal modulation light r_1 with the same modulation frequency which consists of two reflected light r_{11} and r_{12} from the same scene point. r_{11} is the reflection of itself and r_{12} is the interference reflection light from C_2. Similarly, C_2 receives the combined signal r_2 which consists of r_{22} and r_{21}. r_1 and r_2 can be written as

$$\begin{cases} r_1 = r_{11} + r_{12} \\ r_2 = r_{21} + r_{22} \end{cases} \tag{7}$$

r_{11}, r_{12}, r_{21} and r_{22} can be written as

$$r_{ij} = B_{ij} + A_{ij}\cos(2\pi f_0 t + \varphi_{ij}) \tag{8}$$

with i and j being 1 or 2, where the signal travels from illuminator j to sensor i.

It is further illustrated by analyzing r_1. Based on Eq. (7) and Eq. (8), we rewrite r_1 as

$$\begin{aligned} r_1 &= \tilde{B} + A_{11}\cos(2\pi f_0 t + \varphi_{11}) + A_{12}\cos(2\pi f_0 t + \varphi_{12}) \\ &= \tilde{B} + \tilde{A}\cos(2\pi f_0 t + \tilde{\varphi}) \end{aligned} \tag{9}$$

with $\tilde{B} = B_{11} + B_{12}$, where $\tilde{A}$ and $\tilde{\varphi}$ are the resultant amplitude and phase shift sensed by C_1. After 4 steps sampling, the measurement results can be written as

$$\tilde{A} = P_0\sqrt{(A_{11}^2 + A_{12}^2 + 2A_{11}A_{12}cos(\varphi_{11} - \varphi_{12}))}/2 \tag{10}$$

$$\tilde{\varphi} = \arctan\left(\frac{A_{11}\sin\varphi_{11} + A_{12}\sin\varphi_{12}}{A_{11}\cos\varphi_{11} + A_{12}\cos\varphi_{12}}\right) \tag{11}$$

As a consequence, because of the interference components A_{12} and φ_{12}, erroneous amplitude $\tilde{A}$ and phase $\tilde{\varphi}$ are extracted instead of the desired values given by A_{11} and φ_{11}, respectively. So TOF cameras obtain wrong range and amplitude data due to the corrupted sensor pixels with multi-camera interference.

4. MULTI-CAMERA INTERFERENCE CANCELLATION

The goal of multi-camera interference cancellation is to accurately estimate the phase φ (or distance d) and the amplitude A from the interfered light signal. In frequency domain, traditional methods fail to solve the problem with the same modulation frequency. But from the time domain of multiframe, we find an available estimation method based on the statistical property of interference signal.

4.1. Problem analysis

Though the illumination light of TOF camera is a continuous wave, its optical modulation is only active during the sampling time (called as integration time too) and turned off for the intermittent period in each frame. It means that TOF camera completes the whole sampling in one frame, which is independent and unrelated to other frames. In addition, for each pixel, the measured values are independent and identically distributed. So, for N frames, the received signal of camera C_1 can be considered as a sequence $\{r_1(i), i = 1 \cdots N\}$ of independent identically distributed random variables, which can be written as

$$\begin{aligned} r_1(i) &= \tilde{B}(i) + A_{11}(i)\cos[2\pi f_0 t + \varphi_{11}(i)] \\ &+ A_{12}(i)\cos[2\pi f_0 t + \varphi_{12}(i)] \end{aligned} \tag{12}$$

In each frame, every camera completes the independent measurement with different initial phase of illumination light. Thus for multiple frames, the $\varphi_{11}(i)$ is constant and equal to the truth phase φ_{11}, but the interference phase $\varphi_{12}(i)$ is a variable due to the asynchronous operation between two cameras. In addition, for the same point in different frames, the light propagation can be treated as constant in static scenes. So without loss of generality, we assume $\tilde{B}(i)$, $A_{11}(i)$ and $A_{12}(i)$ are unchanged and equal to $\tilde{B}$, A_{11} and A_{12} respectively. So, Eq. (12) can be written as

$$r_1(i) = X_1 + I(t) \tag{13}$$

with noninterference component $X_1 = \tilde{B} + A_{11}\cos(2\pi f_0 t + \varphi_{11})$, interference component $I(t) = A_{12}\cos(2\pi f_0 t + \phi)$, and random phase $\phi = \varphi_{12}(i)$.

4.2. LSENS

In Eq. (13), if X_1 is accurately estimated, the phase and amplitude information can be recovered from the interfered signals $r_1(i)$. The least squares estimation $\hat{X}_1$ of X_1 is an optimal estimation, which equals $E[X_1]$ [12].

$$\hat{X}_1 = E[X_1] = E[r_1(i)] - E[I(t)] \tag{14}$$

The interference item $I(t)$ is a random phase process, where ϕ is the random variable which is uniformly distributed in the interval $(0, 2\pi)$. This is a second order process with mean as

$$E[I(t)] = m_I = \int_0^{2\pi} A_{12}\cos(2\pi f_0 t + \phi)\frac{1}{2\pi}d\phi = 0 \tag{15}$$

and correlation function as

$$R_I(t, t-\tau) = A_{12}\cos(2\pi f_0 t\tau)/2 \tag{16}$$

The random phase process is the wide-sense stationary stochastic process and can be proved to be ergodic [13]. So the interference component $I(t)$ is proved to be an ergodic and wide-sense stationary stochastic process. Thus, the ensemble average of $I(t)$ can be replaced by the time average of any sample function. In addition, in the time domain of multiframe, depth measurement is a discrete process, so Eq. (15) is replaced by $\hat{m}_I \approx \sum_{i=1}^{N} I_i/N$, which will be close to 0 if N is large enough. So, to eliminate the interference, we can compute the mean of $r_1(i)$ and get the estimation as

$$\hat{X}_1 \approx E[r_1(i)] \tag{17}$$

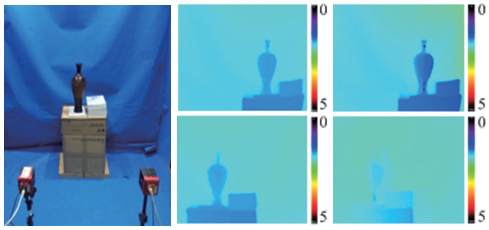
with error $A_{12}\sum_{i=1}^{N}\cos[2\pi f_0 t + \varphi_{12}(i)]/N$. In Eq.(17),when N increases, the error from the interference decreases and the depth and amplitude data recovered from the multi-camera interference are closer to the truth.

5. EXPERIMENTS AND RESULTS

To verify the effectiveness of the proposed method, we perform the experiments on the real scene shown in Fig.3. We using two TOF cameras, MESA Imaging SR4000, to capture the scene with a vase and some books on a box. The vase is placed about 130 cm away from the two cameras which are about 50 cm apart from each other and both work with 30MHz modulation frequency. We capture more than 100 frame depth and amplitude images.

As shown in Fig.3, because of the multi-camera interference problem, the depth images captured by two cameras have obvious errors. As an example, seven frame images captured from left camera are shown in Fig.4. Due to the change of $\varphi_{12}(i)$ in multiple frames, the depth and amplitude of each frame change. The 7th frame depth and amplitude image are similar with the first images.

To evaluate the performance of LSENS, we compute the peak signal to noise ratio (PSNR) with the undisturbed image data (ground truth). As shown in Fig.5(a), with N increasing, the PSNR of the depth images processed with N frame images is improved with fluctuation and converges to a



(a) (b) (c)

Fig. 3. Capturing a scene with two SR4000 cameras operating at 30MHz modulation frequency. (a) Scene for shooting. (b) Depth images captured by left and right TOF cameras without interference. (c) Depth images captured by left and right TOF cameras with multi-camera interference. The colors of depth images represent the ranges from 0 m to 5 m.

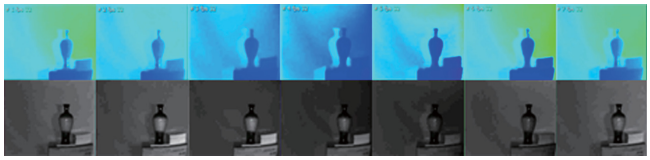
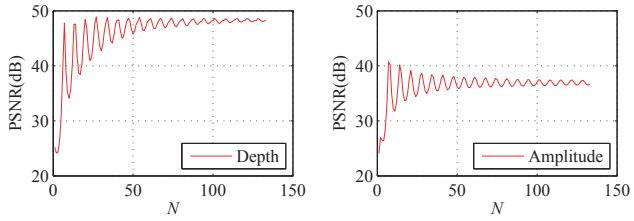


Fig. 4. Multiframe depth and amplitude images captured from left camera with multi-camera interference. The upper images are depth images with interference from image frame 1 to 7. The lower images are 7 frame amplitude images with interference where the intensity represents the amplitude.

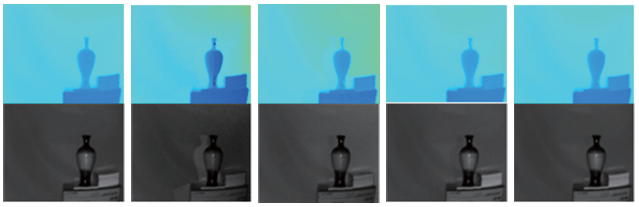
value over 48dB. Because $\varphi_{12}(i)$ is uniformly distributed in the interval $(0, 2\pi)$, when the sampling for multiple frames is close to an approximate cycle, the peak of PSNR appears at $N = 7, 13, 20, 27$ and so on, which are more than 48dB. The PSNR of the processed amplitude images, shown in Fig.5(b), converges to a value over 36dB and has a maximum 40.74 dB at $N = 7$. Fig.6 shows that the qualities of the processed depth and amplitude images are much better than the untreated images.



(a) (b)

Fig. 5. PSNRs of depth images and amplitude images processed by LSENS with N frame images.

The above results are based on the images which are converted from the measured data. To describe the results more accurately, we put the measured data into 3D space to get the point cloud of the 3D scene and compute the root mean square error (RMSE) of the point cloud processed with N frame data. As shown in Fig.7, with N increasing, the RMSE of processed point cloud data decreases significantly, which converges to a value less than 0.01 and has the first valley value 0.01 at $N = 7$. Fig.8 shows that the quality of processed point cloud obtains a great improvement.



(a) (b) (c) (d) (e)

Fig. 6. Depth and amplitude images processed with N frame images in interference case. (a) Ground truth. (b) Depth and amplitude images with interference. (c)-(e) Depth and amplitude images processed with $N = 3, 6, 7$.

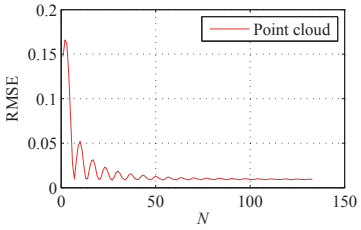
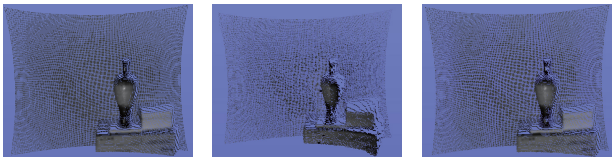


Fig. 7. RMSE of point cloud processed with N frame data.



(a) (b) (c)

Fig. 8. Point cloud of processed data. (a) Ground truth. (b) Point cloud of the data with interference. (c) Point cloud of the data processed with $N = 7$.

Owing to the periodically positive and negative fluctuations of the sine function, even if the frame number N is not large, the processed data are sill close to the truth values. In our experiments, with the frame number $N = 7$, we can get a nice result, shown in Fig.8. For another similar multi-camera cases, the minimum frame number needed by the solution can be decided with practical measurement.

6. CONCLUSION

In this paper, by analyzing the statistical property of multi-camera interference signal under the condition of multiple frames, the interference signal is proved to be an ergodic and wide-sense stationary stochastic process. The LSENS is proposed to remove the multi-camera interference. The results of real experiments prove the method can effectively recover the depth and amplitude information from the severe interference. This approach can be easily applied to commercial TOF cameras without extra customization or upgrade.

7. REFERENCES

[1] A. Kolb, E. Barth, R. Koch, and Ra. Larsen, "Time-of-flight sensors in computer graphics," in *Proc. Eurographics (State-of-the-Art Report)*, 2009, pp. 119–134.

[2] F. Remondino and D. Stoppa, *Tof range-imaging cameras*, Springer, 2013.

[3] J. P. Godbaz, M. J. Cree, and A. A. Dorrington, "Closed-form inverses for the mixed pixel/multipath interference problem in amcw lidar," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2012, pp. 829618–829618.

[4] A. Kirmani, A. Benedetti, and P. A. Chou, "Spumic: Simultaneous phase unwrapping and multipath interference cancellation in time-of-flight cameras using spectral methods," in *Multimedia and Expo (ICME), 2013 IEEE International Conference on*. IEEE, 2013, pp. 1–6.

[5] D. Jiménez, D. Pizarro, M. Mazo, and S. Palazuelos, "Modeling and correction of multipath interference in time of flight cameras," *Image and Vision Computing*, vol. 32, no. 1, pp. 1–13, 2014.

[6] A. Bhandari, A. Kadambi, R. Whyte, C. Barsi, M. Feigin, A. Dorrington, and R. Raskar, "Resolving multipath interference in time-of-flight imaging via modulation frequency diversity and sparse regularization," *Optics letters*, vol. 39, no. 6, pp. 1705–1708, 2014.

[7] B. Buttgen, M. H. A. El Mechat, F. Lustenberger, and P. Seitz, "Pseudonoise optical modulation for real-time 3-d imaging with minimum interference," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 54, no. 10, pp. 2109–2119, 2007.

[8] B. Buttgen and P. Seitz, "Robust optical time-of-flight range imaging based on smart pixel structures," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 55, no. 6, pp. 1512–1525, 2008.

[9] R. Z. Whyte, A. D. Payne, A. A. Dorrington, and M. J. Cree, "Multiple range imaging camera operation with minimal performance impact," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2010, pp. 75380I–75380I.

[10] R. Lange, "3d time-of-flight distance measurement with custom solid-state image sensors in cmos/ccd - technology," *Ph.D. dissertation, Department of Electrical Engineering and Computer Science at University of Siegen*, 2000.

[11] M. Hansard, S. Lee, O. Choi, and R. Horaud, *Time-of-Flight Cameras: Principles, Methods and Applications*, Springer, 2013.

[12] D. P. Bertsekas and J. N. Tsitsiklis, *Introduction to Probability*, Athena Scientific, 2002.

[13] W. A. Gardner, *Introduction to Random Processes: with applications to signals and systems*, McGraw-Hill, 1990.