Hybrid exposure for depth imaging of a
time-of-flight depth sensor

Hyunjung Shim$^{1,3}$ and Seungkyu Lee$^{2,*}$

$^1$ School of Integrated Technology, Yonsei University, South Korea
$^2$ Department of Computer Engineering, Kyung Hee University, South Korea
$^3$kateshim@yonsei.ac.kr
$^*$seungkyu@khu.ac.kr

Abstract: A time-of-flight (ToF) depth sensor produces noisy range data due to scene properties such as surface materials and reflectivity. Sensor measurement frequently includes either a saturated or severely noisy depth and effective depth accuracy is far below its ideal specification. In this paper, we propose a hybrid exposure technique for depth imaging in a ToF sensor so to improve the depth quality. Our method automatically determines an optimal depth for each pixel using two exposure conditions. To show that our algorithm is effective, we compare the proposed algorithm with two conventional methods in qualitative and quantitative manners showing the superior performance of proposed algorithm.

© 2014 Optical Society of America

OCIS codes: (110.0110) Imaging systems; (100.0100) Image processing; (110.6880) Three-dimensional image acquisition; (100.6890) Three-dimensional image processing; (100.2980) Image enhancement; (010.1080) Active or adaptive optics.

References and links

1. P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox, “Rgbd mapping: using depth cameras for dense 3-D modeling of indoor environments,” in RGB-D: Advanced Reasoning with Depth Cameras Workshop in conjunction with RSS, The MIT Press, Cambridge, Massachusetts (2010).
2. S. Foix, G. Alenya, and C. Torras, “Lock-in time-of-flight (ToF) cameras: a survey,” IEEE Sens. J. 11(9), 1917–1926 (2011).
3. A. Colaco, A. Kirmani, N.-W. Gong, T. McGarry, L. Watkins, and V. K. Goyal, “3dim: Compact and Low Power Time-Of-Flight Sensor for 3D Capture Using Parametric Signal Processing,” Proc. Int. Image Sensor Workshop, 349-352 (2013).
4. S. Schuon, C. Theobalt, J. Davis and S. Thrun, “High-quality scanning using time-offlight depth superresolution,” IEEE Conference on Computer Vision and Pattern Recognition Workshop, 1–7 (2008).
5. H. Shim and S. Lee, “Performance evaluation of time-of-flight and structured light depth sensors in radiometric/geometric variations,” Opt. Eng. 51(1), 94401–94414 (2012).
6. S. B. Gokturk, H. Yalcin and C. Bamji, “A Time-Of-Flight Depth Sensor - System Description, Issues and Solutions,” IEEE Conference on Computer Vision and Pattern Recognition Workshop, 35–44 (2004).
7. V. Brajovic and T. Kanade, “A Sorting Image Sensor: An Example of Massively Parallel Intensity-to-Time Processing for Low-Latency Computational Sensors,” Proceedings of the IEEE International Conference on Robotics and Automation, 2, 1638–1643, (1996).
8. S. K. Nayar and V. Branzoi, “Adaptive dynamic range imaging: optical control of pixel exposures over space and time,” IEEE International Conference on Computer Vision, 1168–1175 (2003).
9. M. Aggarwal and N. Ahuja, “High Dynamic Range Panoramic Imaging,” International Conference on Computer Vision, 2–9 (2001).
10. O. Yadid-Pecht and E.R. Fossum, “Wide intrascene dynamic range CMOS APS using dual sampling,” IEEE Trans. Electron Dev. 44(10), 1721–1723 (1997).
11. P.Debevec and J. Malik “Recovering high dynamic range radiance maps from photographs,” in ACM SIGGRAPH classes (ACM Press, 2008, 31).
12. T. Mertens and J. Kautz and F. Van Reeth “Exposure Fusion,” IEEE Pacific Conference on Computer Graphics and Applications, 382–390 (2007).

Received 26 Mar 2014; revised 1 May 2014; accepted 3 May 2014; published 27 May 2014

(C) 2014 OSA
1. Introduction

Time-of-flight (ToF) depth sensors provide three-dimensional information of an object. If we can obtain a range data that is invariant to the illumination and texture of an object, it will be a great benefit in addressing important research problems, such as object detection, tracking, and recognition. Thanks to the nice properties like real-time acquisition, affordable price and portability, consumer depth sensors have been successfully adopted in many applications, including 3D scene reconstruction[1], augmented reality, gesture recognition, tracking, pedestrian detection and game.

Despite its success in those applications, the use of depth sensor is still restricted by its inherent limitations. It has the low precision (low signal-to-noise ratio)[2, 3], low resolutions[4] and errors caused by geometric and radiometric variations[5]. Fortunately, the precision and resolution problems are getting mitigated rapidly powered by the advancement in sensing technology. However, errors due to geometric and radiometric variations still remain as an open problem.

Both issues arise because the conventional sensors have a limited dynamic range due to their practical implementation issues as addressed in [6]. Existing ToF depth sensor emits the IR signals to target objects and measures the phase delay of reflected IR signals at every pixel to obtain the range data. Due to the limited dynamic range of sensor, the excessive amount of reflected lights yields the saturation while the lack of reflected lights leads to the low signal-to-noise ratio (SNR) in measurement. This phenomenon occurs when the target object is geometrically quite close to or far from the sensor. This also appears when the target object is specular. The specular object reflects the incoming light toward a narrow range of viewing direction. Therefore, depending on the relative viewing direction of sensor, the recorded range data presents either saturation or noise.

Our work aims to enhance the quality of range data using two exposure conditions. A key idea is to characterize the saturation and noise by an effective imaging architecture using two exposures and determine the optimal depth per pixel in the time-of-flight depth sensor. As a result, we improve the quality of range data. Utilizing variable exposures is popularly adopted in the study of High Dynamic Range (HDR) color imaging. Existing work can be categorized by a single capture and multiple capture technique. For a single capture technique, a idea of controllable capacitor[7], attenuator[8], motion[9], exposure[10] are employed in last decade. Brajovic and Kanade[7] introduce a novel imaging sensor for HDR imaging, implementing controllable capacitors to adjust the level of incoming light. Nayar and Branzoi[8] introduce a controllable liquid crystal light modulator to attenuate the incoming light per pixel and change the attenuation factor per pixel according to the scene radiance. A multiple capture technique[11, 12] acquires multiple images at different exposures and fuse them to obtain a single high quality image. There are a couple of advantages and disadvantages in each category. Single capture techniques are more desirable for dynamic scene (except[9]) while the quality of output is significantly worse than multiple captures[13]. Multiple capture techniques achieve the better quality while they are limited to static scene. The advantage of our method is to maintain the strength of both sides, achieving the substantial quality improvement and retaining the capa-
bility of real time capture. In fact, although existing approaches have made the meaningful technical contributions in high quality imaging, they cannot be directly employed to our problem. It is because the color image is interpreted as the collection of photons while the range data is derived by the profile of photons over time. This leads us to reformulate the problem, characterizing the erroneous measurements to improve the quality of range data. In our problem, we use two depth measurements to examine their variation in the frequency domain, characterize their variation as the noise component and filter them out to obtain the robust depth.

Recently, Hahne and Alexa [14] introduce a new method to fuse four depth maps for obtaining the high quality depth data. Using corresponding amplitude images, they extract four different confidence measures and use them to compute the weighting map for each depth map. Then, they determine the final high quality depth map by the weighted sum of multiple depth maps corresponding to various exposures. Since they assign a weight per pixel, they can combine depth values differently according to pixel. However, their method is not reliable to handle highly saturated and noisy depth maps, such as the depth map with the full distance range or specular object. It is because the amplitude image is also the low dynamic range data presenting the severe distortion with the high dynamic range scene. Under extreme conditions, referencing unreliable amplitude images misleads the confidence of depth value. Indeed, the use of amplitude image can add the distortion to solve the same type of distortion.

In this paper, we only refer two range measurements recorded at a short and long exposure condition. These two range measurements show two distinct characteristics. A range data from a short exposure is effective to capture specular and highly reflective surfaces while it presents severe noise. One from a long exposure is robust to sensor noise while it includes the saturation error on highly reflective surfaces. The saturated pixel can be identified by a simple thresholding. One possible way to handle the saturated pixel is to replace it with depth values recorded at a different exposure. Yet, depth from different exposures present a different level of noise. Consequently, direct depth fusion yields visible boundary and artifacts produced by inhomogeneous noise after fusion. Using our framework, we characterize the noise components by comparing range data at two different exposures. Subsiding the noise components, we can obtain the reliable depth, which is robust against the noise and saturation.

We believe that our solution is effective to capture the specular object using a time-of-flight sensor. Also, our framework is practical to implement because it builds upon the conventional sensor without installing the pixel-wise controller. Although increasing the number of exposures generally help improving the output quality, our method provides the sufficient improvement even using two exposures. Therefore, our work is suitable for real-time applications. In remaining sections, we first analyze the range data recorded at many exposure conditions and use them to optimize the robust depth for a static scene. Then, we demonstrate the proposed approach for real-time applications. Finally, to evaluate the performance of proposed approach, we compare our results with an ideal single exposure[15] and [14]. The extensive experimental results will justify the strength and effectiveness of our work.

2. Depth profile upon variable exposures

In this section, we analyze the depth profile recorded at multiple exposure conditions and show that it contains sufficient information for a robust depth restoration. By investigating the depth profile, we select the most reliable exposure condition per pixel. Although our proposed method uses only two exposures in the depth profile, we first restore an optimal depth using $N$ exposure conditions, which is regarded as the upper-bound performance of multiple exposure-based techniques.

Suppose that we record $N$ depth measurements in varying exposure conditions. From the set of measurements, $N$ different range values are obtained for every pixel. By concatenating...
these measurements, we build a $N$ dimensional depth exposure profile (DEP) of each pixel. The DEP of an erroneous pixel exhibits either saturation or severe noise pattern. Figure 1(a) shows the DEP of saturated pixel, more specifically the pixel for specular highlights. Although this includes some invalid measures (e.g. out-of-range) for a long exposure, the sample for a short exposure still provides a reliable and valid depth value. The noisy DEP is depicted in Fig. 1(b). This appears when the emitted light seldom returns to the sensor. In this circumstance, although the noise level becomes severe for the short exposure, the observation in the long exposure is effective to determine the depth value. These observations exactly match with our hypothesis as well as the analysis from the previous work [5].

Since we have enough samples to analyze the depth profile, we can selectively leverage them to reach an optimal depth value. That is, we employ the local regression model to represent the DEP. Because the local regression technique estimates an unknown function well without any prior, it is suitable to process any DEP at an arbitrary unknown pixel. Suppose that, $t$ indicates the exposure time, $y_t$ is the noisy sample of DEP at $t$, $x$ is an ideal robust depth, $s_t$ is 0 for the saturated pixel at $t$ (1 for the ordinary pixel) and $\varepsilon_t$ denotes a depth noise. As addressed in [15], the depth noise is increased upon the decrease of exposure time. Hence, we formulate a depth noise as a function of exposure time. Finally, their relationship can be stated by

$$y_t = s_t(x + \varepsilon_t). \tag{1}$$

Observing many samples of measurements, we find that an adaptive local model is sufficient to approximate $x$, where $x = d, t \in [t_0 - H, t_0 + H]$. Then, the final depth $x$ can be computed for $[d, H, t_0]$ that fits our DEP the best and it can be expressed as,

$$[d, H, t_0] = \arg\min_{d, t_0} \sum_{d=t_0-H}^{t_0+H} s_t R(t, t_0) ||y_t - d||^2. \tag{2}$$

$s_t$ is simply computed by checking if $y_t$ is within the operating range of depth sensor. We apply a quasi-Newton method to iteratively determine $H$ by inversely proportional to the residuals in Eq. (2). Although a quasi-Newton method finds a local minima of objective function, it is effective to our problem since $s_t$ removes some invalid samples corrupted by saturation. In our implementation, we use a cm unit for both $y_t$ and $d$ and a ms unite for $t$. We set an initial
as the largest integer value smaller than 0.02N and iteratively update it by the largest integer less than 0.02N and includes more samples if the average distance error is less than 1 cm. Such an adaptive scheme helps our estimates more reliable. Also, we apply a regularization function \( R(t, t_0) \) to introduce two constraints: assigning the high confidence to close sample from \( t_0 \) (first term) and enforcing the selected exposure as large as possible (second term).

\[
R(t, t_0) = \left| \frac{t - t_0}{H} \right| \times \frac{1}{t_0^\alpha}. \tag{3}
\]

The first regularization term is a simple symmetric decreasing function of their distance from the center of the neighborhood, which is commonly used in the local regression method and assigns larger weights for center samples than the rest. The second term is inspired by [15] in that the sample with a long exposure is more reliable than one with a short exposure in general. We find that \( \alpha = 2 \) works well in practice.

Afterwards, we solve Eq. (2) and determine \( d \) for the final depth value. This is repeated for every pixel. Later, we apply this scheme to compare with our solution using two exposures.

3. Hybrid exposure for depth recovery

The proposed algorithm combines two exposures to handle both saturation and noise. The depth noise in Eq. (1) is a function of a shot noise. Scene context (e.g. reflectivity or intensity) affects the signal level and change the signal-to-noise ratio. A variety of studies shows that the shot noise is a stationary process, however the observed noise in ToF depth image shows non-stationary characteristic [16]. Therefore, it is important to learn the noise characteristics adaptive to our observation. Figure 2 visualizes the spatial frequency responses of multiple depth maps at different exposures. The ringing artifact gets severe in the frequency responses of longer exposures because the saturated regions is increased according to exposures. In these examples, we observe that the coefficients of high frequency components increase as we decrease exposure, which correspond to the increase of shot noise. Based on our two observations in long exposure \( (y_{ti}) \) and shot exposure \( (y_{ts}) \), we characterize the increment of noise by different exposures.

More specifically, we identify the noise difference in frequency domain and restore accurate depth value in short exposure up to the level of accuracy in long exposure. From Eq. (1), noise terms of long and short exposures in a non-saturated pixel (i.e. \( \forall s_t = 1 \)) are written as follows.

\[
\epsilon_{ts} = y_{ts} - x, \quad \epsilon_{tl} = y_{tl} - x. \tag{4}
\]

The noise difference between a pair of exposures is denoted by \( \epsilon_\delta \) and defined as follows.

\[
\epsilon_\delta = \epsilon_{ts} - \epsilon_{tl} = y_{ts} - y_{tl}, \quad \forall s_t = 1. \tag{5}
\]

\( y \) is a matrix of all depth measurements at a single exposure.

Based on Eq. (5), the frequency components of noise difference is identified by a pair of range data. We eliminate \( \epsilon_\delta \) by designing the low pass filter with a cut-off frequency of \( \omega_c \). The size of saturation blobs is used to guide the cut-off frequency for \( \omega_c \). The scheme to determine \( \omega_c \) is simple as follows.

1. For horizontal components, determine the horizontal size of saturation blobs and compute the maximum difference of blob size, \( BW_{max} \).
2. Choose $\pi(W - BW_{\text{max}})/W$ as $\omega_w^W$ when $W$ is the horizontal pixel resolution.

3. For vertical components, determine the vertical size of saturation blobs and compute the maximum difference of blob size, $BH_{\text{max}}$.

4. Choose $\pi(H - BH_{\text{max}})/H$ as $\omega_h^H$ when $H$ is the vertical pixel resolution.

Such a method suppresses the noise components with a spatial period smaller than $BW_{\text{max}}$ in width and $BH_{\text{max}}$ in height. This is effective to remove the noise pattern with a moderate size. It is important because the similar noise often appears in few pixels simultaneously. Notice that the high dynamic range of scene radiance causes the noise and the local neighborhood exhibits the similar surface characteristics. Our method explicitly considers the size of noise pattern to design a noise filter.

From the cut-off frequency for both horizontal and vertical axis, we construct the low pass filter. We apply an anisotropic 2D Gaussian filter with a cut-off frequency of $\omega_w^W$ and $\omega_h^H$. It generates a smoothly varying curve, which helps mitigating the ripple artifacts in depth restoration. Denote that $\hat{\mathbf{f}}$ represents the FFT(fast Fourier transform), $\mathbf{L}$ is the FFT of low pass filter, $\hat{\mathbf{y}}_{tl}$ is the FFT of $y_{tl}$ and $\cdot$ is a pixel-wise multiplication. To apply this filter on noisy range data, we compute

$$\hat{\mathbf{x}} = \hat{\mathbf{f}}^{-1}(\mathbf{L} \cdot \hat{\mathbf{y}}_{tl}). \quad (6)$$

Our method possesses two important properties. First, we remove the noise by examining its characteristics dependent on input data. Second, the proposed algorithm improves the quality of range data up to a level of noise at the long exposure condition. In the following section, we demonstrate the performance of the proposed algorithm by various experimental study.

4. Experimental results

For the implementation, we use MESA SR4000 of which exposure varies from 0.3 to 25.8 ms with the resolution of 0.1 ms and operating range is between 0.1 and 5 meter. Throughout all
the experiments, we set 0.3 ms and 25.8 ms for two exposure conditions so to manage real time applications. Since the biggest advantage of ToF depth sensor is its ability on real time acquisition, it is important to retain such a capability. We demonstrate that our framework is effective even using two exposure conditions, suitable for a dynamic scene.

For the evaluation, we compare our results with the conventional single exposure [15] and the amplitude based exposure fusion method [14]. For a fair comparison with [14], we use the code from the authors, the best parameter setting and four exposure conditions \{0.3m, 8.8m, 17.3m, 25.8m\}. In addition, we apply our scheme described in Sec. 2 with 75 exposure conditions. We regard it as the reference for our achievement, the upper-bound performance by employing multiple exposure conditions. Experimental evaluations are conducted on ten challenging test scenes, which are recorded at common office and house environments. Our test scenes are challenging as they cover the full range of distance of the depth sensor with highly specular objects.

Figure 3 visualizes the living room with an LCD TV and each column corresponds to the different method. Using the conventional single exposure, we find that the center region of the display presents data loss due to saturation. This problem is alleviated by Hahne and Alexa [14] in that they fill the missing data with some valid depth values from low exposure sample. However, as shown in the profile view, overall scene becomes more noisy because of the choice of non-optimal exposure sample for other pixels. Both of our results correctly reconstruct the missing region and even eliminate overall scene noise producing the visually pleasing result. Comparing two exposures in third column and 75 exposures in fourth column, the results of 75 exposures present less noise but more reliable performance as expected. Still, our algorithm using two exposures sufficiently improves the quality of range data, quite comparable with that of 75 exposures. Figure 4 demonstrates the experimental results of various challenging scenes. A living room in Fig. 3 and the kitchen in Fig. 4(c) present the large intensity and reflectivity variation. Please refer Fig. 5 for their close-up views. Figure 5 highlights that the proposed algorithms improves the quality for saturated pixels significantly. The office table in Fig. 4(b) and the dining table in Fig. 4(b) span the full depth range with various radiometric properties. Especially, they have highly specular objects: the plastic teapot and flowerpot in Fig. 4(a) and shiny porcelain and metal bowl in Fig. 4(b). As seen from the results, the proposed algorithm enhance the quality of 3D scene with the high precision. The same observation consistently holds in other experiments.

For quantitative evaluation, we measure 1) the temporal repeatability and 2) the spatial repeatability by recording known planar objects and calculating the residuals of plane fitting.
First, we capture each scene 1000 times and use them to measure the standard deviation of respective results. Figure 6 illustrates the average standard deviation of three different algorithms across the different scene. The overall standard deviations are 2.10 cm for the single exposure, 4.16 cm for four exposures [14], 1.11 cm for the proposed algorithm using two exposures and 0.43 cm for ours using 75 exposures. In addition to the overall precision, our algorithm is robust to changes in target scene. Note that our test scenes include highly specular objects presenting the over-saturation both in amplitude and depth map. As a result, Hahne and Alexa [14] often underperforms because they use the corrupted amplitude map to enhance the depth quality. This is why they show limitation in temporal repeatability.

To test the spatial repeatability, we capture the depth maps of a white board placed at five different distances: each corresponding to approximately 0.5, 1.5, 2.5, 3.5, 4.5 meters. Then, we apply three different methods to process them and fit a plane to each result by the least-squares method. We compute an average residuals and use it to measure the spatial repeatability. Figure 7 compares the spatial repeatability of different algorithms. The overall residuals are 0.502 cm for the single exposure, 0.344 cm for [14] using four exposure conditions, 0.198 cm for the proposed algorithm using two exposure conditions and 0.148 cm for ours using 75 exposure conditions. Figure 8 visualizes the results for plane at 4.5 meter in 3D. From Fig. 7, we find that Hahne and Alexa performs comparable except for the plane at 0.5 m. It is because the white board scene does not include highly specular objects. However, the plane object at 0.5 m presents the saturation caused by its near distance so that the overall precision drops substantially. This is analogous to the observation in Fig. 3, sacrificing the precision to handle the strong saturation. Similar to the temporal repeatability evaluation, the proposed algorithm
shows the robust performance to changes in the scene.

Although the proposed algorithm generally improves the quality of range data, it is not effective to correct the errors produced by the interreflection. The interreflection, also referred to as the multipath problem, occurs at a corner of a scene or a concave object. This introduces the additional light reflection to the sensor and becomes another source of errors. Currently, our method ignores such an issue. Hence, the proposed algorithm is effective for convex objects without the presence of interreflection.

5. Conclusion

In this paper, we introduced an effective solution to improve the quality of depth map of a Time-of-flight depth sensor. Our algorithm works with any type of ToF sensors, which exposure setting is controllable. Because the proposed algorithm uses only two exposures conditions, it is appropriate for real time applications. Based on the extensive experiments and evaluation, we have shown that the proposed algorithm performs better than the conventional algorithms in terms of temporal and spatial repeatability even using much less input measurements. More importantly, the proposed algorithm generates the visually pleasing results, clearly outperforming existing methods. The proposed algorithm is attractive due to not only its outstanding perfor-
Fig. 7. Residuals of plane fitting. Overall residuals are 0.506 cm (single), 0.34 cm ([14]) and 0.19 cm (proposed).

Fig. 8. Spatial repeatability using plane fitting.

mance but also its robustness to the changes of scene. Moreover, our solution reconstructs the specular object using a single time-of-flight sensor. Thanks to its effectiveness and robustness, we expect that our method can improve the performance of existing work using ToF sensors and further advance its potential applications.

Acknowledgment

We thank Uwe Hahne for his kind courtesy of providing a source code and the detailed explanations. This research was supported by the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the IT Consilience Creative Program (NIPA-2014-H0201-14-1001) supervised by the NIPA(National IT Industry Promotion Agency) and also the Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Science, ICT and Future Planning (2013062644).