High-dynamic-range Imaging for Depth Maps

Hanseul Jun
Stanford University
hanseul@stanford.edu

Abstract

In this paper, I propose a method which applies the fundamental idea and technique of high-dynamic-range (HDR) imaging [5] to depth maps obtained from time-of-flight depth sensors. HDR imaging creates a single high-quality output from a number of photographs in different settings as its input. Similarly, my technique uses a set of depth maps, which are obtained from different settings (i.e. frequency of the time-of-flight camera [8]), to create a single high-quality depth map. From the fact that differently ranged depth maps have different sweet spots, this method results in having a depth map that is both detailed in short distances and robust in long distances.

1. Introduction

Depth map, a form to contain 3D space information, has a massive variety of applications. One of them is the field of augmented reality which has a high demand of 3D information. In this field, depth maps are the de facto standard form of 3D information since there are decent commercial devices that can create depth maps. Among them, the ones under the greatest interest are time-of-flight depth sensors. These devices with decent performances are cheap, light, and easy to use. As a result, virtually every augmented reality headsets that does tracking has a time-of-flight depth sensor on it. For example, a Hololens[1]—currently the most popular augmented reality headset—has four time-of-flight depth sensors working together for simultaneous localization and mapping [9]. Though there are four of them, thanks to the light weighted small depth sensors, Hololens is still a mobile wireless device.

Depth maps look similar to RGB images in the sense that they are 2D arrays. However, the process to create them is widely different. While RGB images can be obtained by filtering all light except some in ranges of frequency and integrating them, this is not the case for depth maps. To measure depth using electro-magnetic signals, which moves at the speed of light, the measurement device should handle the level of time that matches distances divided by the speed of light. This makes the construction of precise depth maps a significantly difficult task. An eminent technique to overcome this problem is to measure time-of-flight indirectly. By signal processing techniques, the measurement of time-of-flight can be done without directly handling the level of time mentioned above. However, this technique inevitably results in having a new problem which is phase unwrapping. The proposed technique is basically to solve this phase unwrapping problem, adopting the technique that was successful for solving a problem of RGB images.

[1] https://www.microsoft.com/en-us/hololens

2. Related Work

A significant amount of work has been done for the phase unwrapping problem with currently available hardwares. While many including me proposed to use multiple frequencies to solve this problem, some others used a single frequency to measure depth, which inherently needs an assumption to solve the problem.

Choi *et al*. [2] tried to unfold—extend the range beyond its conventional limit—the range of depth cameras. They used a model based on the Markov random field which made a significant achievement in their results. They reported that their average success rate is 92.4%. Crabb and Manduchi [3, 4] have made efforts to overcome the limit of range the depth cameras only with a single frequency. Stating that the unwrapping problem is not unique to the field of time-of-flight depth imaging, they tend to adopt techniques from other fields and [3] is a probabilistic interpretation toward the phase of depth camera signals, where folding happens. After overcoming [2], they made a step further in [4]. In this paper, they refer "the use of multiple frequencies" as a de facto solution. The papers claimed that the usage of multiple frequencies [7, 6, 1] achieved better precision utilizing the redundancy from the range of frequencies. This result matches one of my motivating conjectures that usage of multiple frequencies will improve the quality of depth maps.

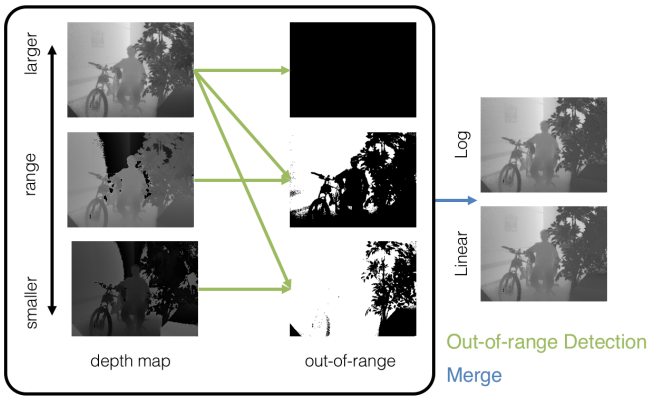

Figure 1. Overview of the method.

3. Method

This method is based on the concept of HDR [5]. Therefore, the mathematical technique stems from the mathematical model of it. However, I adopted a different model in addition to the log-scale model, since the reason of using log-scaled model comes from the perceptual nature of human visual systems, which does not apply to depth sensors. In the case of depth maps, usually, the purpose is not to appreciate the outcome but to run another algorithm which needs depth information to accomplish its task. Therefore, the main concern in building the model may shift to the nature of error terms. Since our method is mainly for time-of-flight cameras which suffer from the phase unwrapping problem, one of my approach is to use the linear-scale model. I still adopted the log-scaled model from a semantical concern. For example, the difference between 1 m to 2 m is not the same to 10 m to 11 m in a viewpoint. When this resonates with the following program using the depth maps, then using the log-scale model might be a better choice.

3.1. Algorithm

This method is based on the idea of HDR, which is basically cherry-picking information from measurements in different settings. For this purpose, ideally, the depth map with the largest range should be large enough to measure everywhere in the scene. If the largest range is not large enough to measure the whole scene, no matter how other depth maps with different range are created, information loss is unavoidable.

My technique detects out-of-range pixels from the depth maps, then merges them into one. The details of the process in Figure 1 will be discussed in Section 3.1.1 and 3.1.2.

3.1.1 Out-of-range Detection

This step works with an assumption that the depth map with the largest range does not have pixels out-of-range. In other words, the method assumes there is a depth map that has the range that covers the whole scene. When the distance in the depth map with the largest range is $d_{largest}$, with a threshold θ, we consider a pixel of a depth map whose range is r_i is out-of-range if $d_{largest} > \theta \times r_i$.

3.1.2 Merge

Using the depth maps and the outputs from the out-of-range detection, the step merges the depth maps into one, ignoring the pixels that are out-of-range. The mathematical details of this step will be discussed in Section 3.2.

3.2. Mathematical Technique

Since [5] uses the log-scale model and I follow a similar approach, I would first describe this method in terms of the log-scale model, then the linear-scale model, which is a modification of it.

3.2.1 Log-scale model

The following equations are to obtain the output ($\hat{X}$), where D_i is the normalized input of a depth map and r_i is the range of a depth map.

$$\hat{X} = \min_X O \tag{1}$$

$$O = \sum_i w_i \left(\log\left(D_i\right) - \log\left(\frac{X}{r_i}\right) \right)^2 \tag{2}$$

$$\text{where } w_i = \exp\left(-4 \frac{(D_i - 0.5)^2}{0.5^2} \right). \tag{3}$$

From here, I can derive

$$\frac{\partial O}{\partial \log(X)} = \sum_i w_i \left(-2\log(X) + 2\log(D_i) + 2\log(r_i)\right) \tag{4}$$

With the following convex optimization technique,

$$\left.\frac{\partial O}{\partial \log(X)}\right|_{X=\hat{X}} = 0 \tag{5}$$

I have

$$\hat{X} = \exp\left(\frac{\sum_i w_i (\log(D_i) + \log(r_i))}{\sum_i w_i} \right). \tag{6}$$

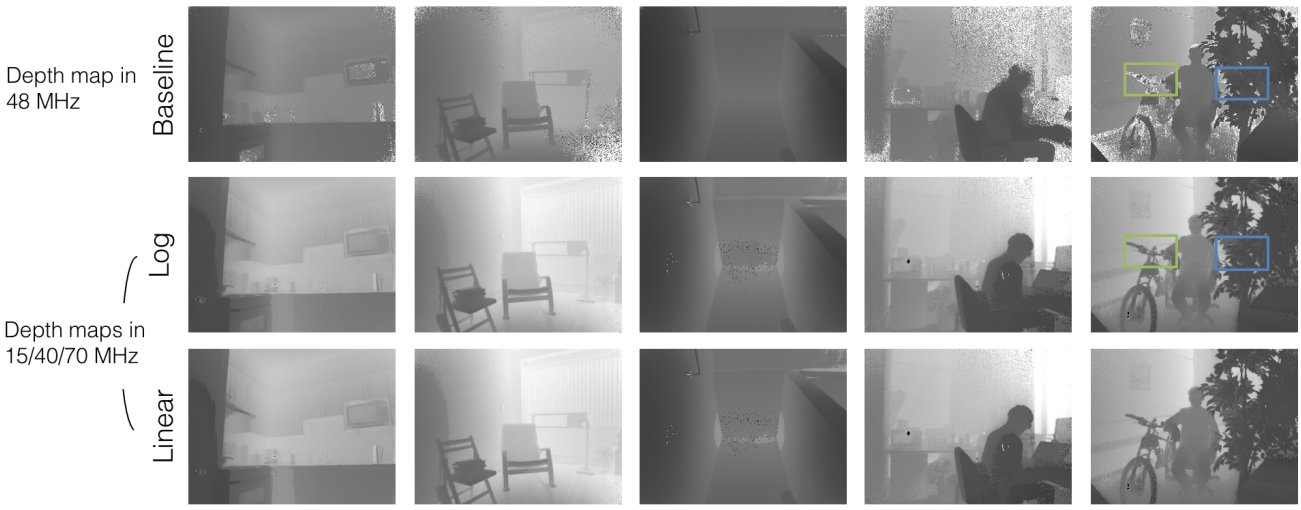

Figure 2. Comparison with the baseline.

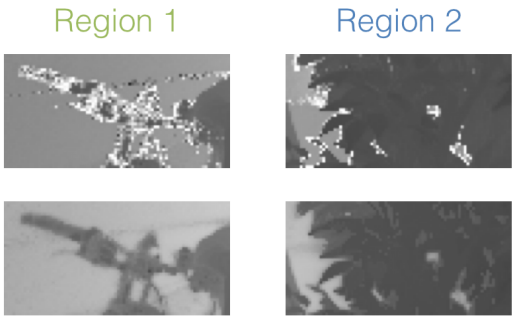

Figure 3. Details from Figure 2.

3.2.2 Linear-scale model

The linear-scale model can be preferred when the error term of depth sensors is the main concern. Using equation 1 and 3, but replacing Equation 2 to the following equation,

$$O = \sum_i w_i \left(D_i - \frac{X}{r_i} \right)^2 . \tag{7}$$

and using a similar convex optimization technique,

$$\left. \frac{\partial O}{\partial X} \right|_{X=\hat{X}} = 0 \tag{8}$$

The result corresponding to Equation 6 is

$$\hat{X} = \frac{\sum_i \frac{w_i D_i}{r_i}}{\sum_i \frac{w_i}{r_i^2}}. \tag{9}$$

4. Results

A dataset with depth maps in 320×240 resolution with 3 different frequencies (15, 40, 70 MHz) was used for the analysis. The baseline was obtained from the dataset with 48 MHz frequency, using the default algorithm of OPT8241[2]. In the preprocess stage, 5 times of measurements in each settings were averaged.

4.1. Qualitative Evaluation

From Figure 2, the images using the proposed methods is more robust in further distances, revealing the proposed method's strength of having an input from a larger range (i.e. 15 MHz). Investigating the details, in region 1 of Figure 3, compared to the baseline method, the proposed method is better for glittery objects and areas with large gradients. Also, in region 2, the depth map from the proposed method contains more details of the leaves, which are in the range of the depth map with 70 MHz frequency, which has a smaller range than the one with 48 MHz.

With these findings, I can say that, as HDR imaging contains information from the brightest and the darkest part of the images, this method contains the details of the closest objects while being robust to the objects in distance at the same time.

4.2. Quantitative Evaluation

Since I could not obtain a pair of ground truth and measurements, simulations were conducted to analyze the proposed method quantitatively. I used the depth maps from the default algorithm of OPT8241 in 48 MHz frequency as

[2]http://www.ti.com/product/OPT8241/technicaldocuments

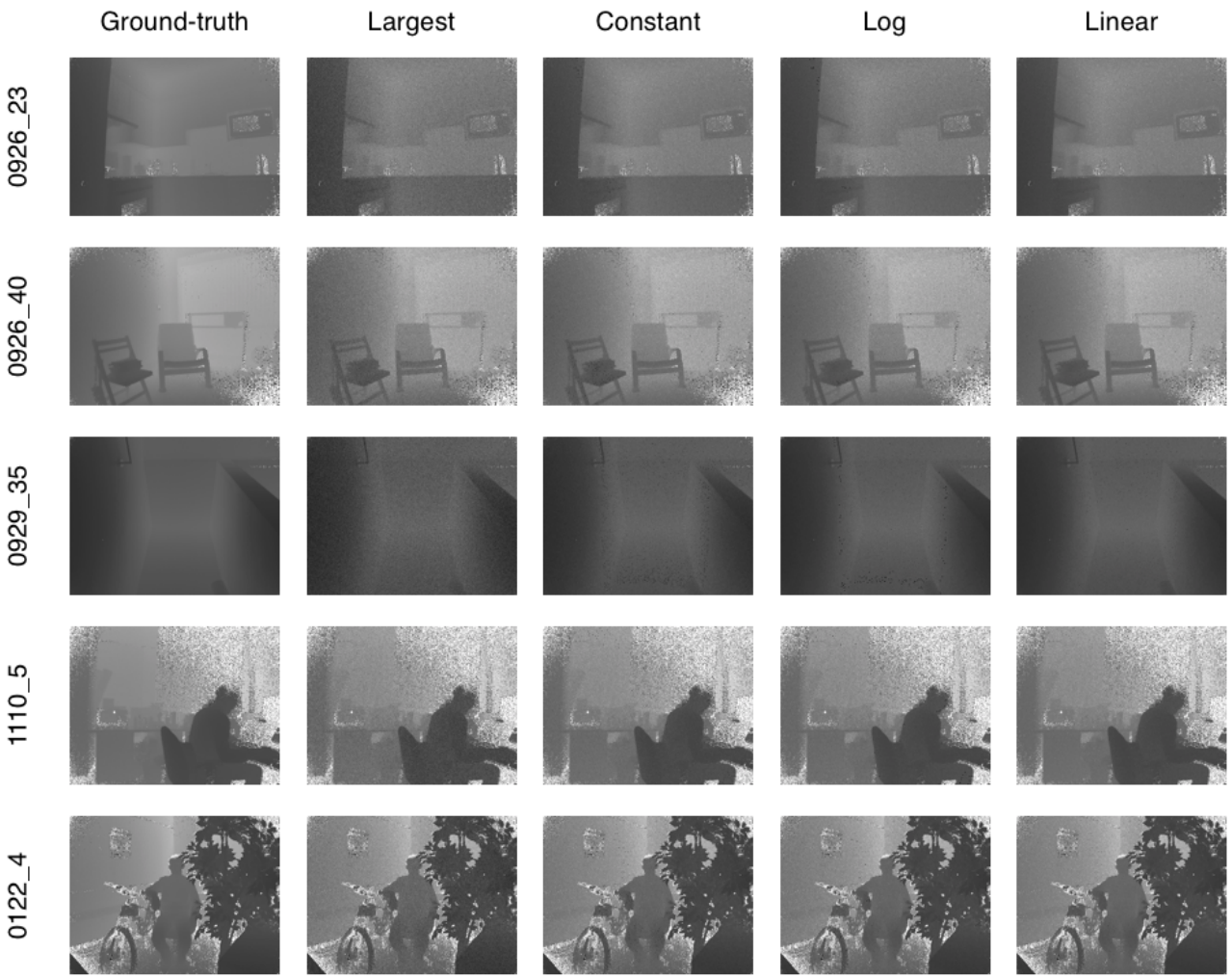

Figure 4. Outputs from the simulation.

Type	Largest	Constant	Log	Linear
0926_23	13.942	17.522	16.730	18.781
0926_40	13.941	15.058	15.001	15.810
0929_35	13.993	17.733	18.656	22.526
1110_5	12.940	14.047	13.840	14.656
0122_4	13.788	16.080	15.976	17.123

Table 1. PSNR values from the simulations.

the pseudo ground-truth. To provide a baseline, I have used two different methods. The first is simply using the depth map with the largest range without any modifications. The second is to use a model without having different weight values (w_i). I added all in-range pixels and merged them by averaging the values. The corresponding formula is

$$\hat{X} = \frac{\sum_i D_i r_i}{\sum_i 1}. \tag{10}$$

To simulate real measurements, Gaussian error terms with standard deviations of (range)$\times 0.02$ were added to each of the depth maps in different frequencies to obtain the results in Table 1.

Apparently, the linear-scale model, which was adopted to reduce the error term was the best in PSNR values. Log-scale model, though it is following the tradition of HDR imaging and concerned toward the semantical meaning of depth, is almost at the same level of the constant-model except for the case of 0929_35. Using the largest ranged depth map was the worst, which is a trivial outcome.

5. Discussion

The application of the idea of HDR imaging to depth maps was at least partially successful. Though I may not be able to call it a state-of-art method, it gathers information from each settings, constructing a single depth map that has the strengths of all inputs. In the qualitative results, the proposed method successfully collected the details of the short ranged depth map while still being robust to the large distances. In the quantitative results, the method, especially the one based on the linear-scale model, increased the PSNR value.

5.1. Future Work

Due to the lack of data, I was not able to test the algorithm using a larger number of settings (i.e. more than 10). Since this technique would only benefit from that change, it might be a way to empower this technique. However, I am not sure how much the advantage would be. Also, using other different settings, not only different frequencies of the time-of-flight camera, should be tested based on the idea of cherry-picking.

6. Acknowledgments

I would like to acknowledge the support of David Lindell for mentoring this project and Shuochen Su for giving advices. Especially, the dataset from Su was essential for the analysis of the proposed method.

References

[1] O. Choi, S. Lee, and H. Lim. Interframe consistent multifrequency phase unwrapping for time-of-flight cameras. *Optical Engineering*, 52(5):057005, 2013.

[2] O. Choi, H. Lim, B. Kang, Y. S. Kim, K. Lee, J. D. Kim, and C.-Y. Kim. Range unfolding for time-of-flight depth cameras. In *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pages 4189–4192. IEEE, 2010.

[3] R. Crabb and R. Manduchi. Probabilistic phase unwrapping for single-frequency time-of-flight range cameras. In *3D Vision (3DV), 2014 2nd International Conference on*, volume 1, pages 577–584. IEEE, 2014.

[4] R. Crabb and R. Manduchi. Fast single-frequency time-of-flight range imaging. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 58–65, 2015.

[5] P. E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 369–378. ACM Press/Addison-Wesley Publishing Co., 1997.

[6] D. Droeschel, D. Holz, and S. Behnke. Multi-frequency phase unwrapping for time-of-flight cameras. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pages 1463–1469. IEEE, 2010.

[7] D. Falie and V. Buzuloiu. Wide range time of flight camera for outdoor surveillance. In *Microwaves, Radar and Remote Sensing Symposium, 2008. MRRS 2008*, pages 79–82. IEEE, 2008.

[8] R. Lange and P. Seitz. Solid-state time-of-flight range camera. *IEEE Journal of quantum electronics*, 37(3):390–397, 2001.

[9] S. Thrun and J. J. Leonard. Simultaneous localization and mapping. In *Springer handbook of robotics*, pages 871–889. Springer, 2008.