RANGE UNFOLDING FOR TIME-OF-FLIGHT DEPTH CAMERAS

Ouk Choi, Hwasup Lim, Byongmin Kang, Yong Sun Kim, Keechang Lee, James D. K. Kim, and Chang-Yeong Kim

Samsung Advanced Institute of Technology, Republic of Korea

ABSTRACT

Time-of-Flight depth cameras provide a direct way to acquire range images, using the phase delay of the incoming reflected signal with respect to the emitted signal. These cameras, however, have a challenging problem called range folding, which occurs due to the modular error in phase delay—ranges are modulo the maximum range. To our best knowledge, we exploit the first approach to estimate the number of mods at each pixel from only a single range image. The estimation is recasted into an optimization problem in the Markov random field framework, where the number of mods is considered as a label. The actual range is then recovered using the optimal number of mods at each pixel, so-named range unfolding.

As demonstrated in the experiments with various range images of real scenes, the proposed method accurately determines the number of mods. In result, the maximum range is practically extended at least twice of that specified by the modulation frequency.

Index Terms—Time-of-Flight depth camera, range folding.

1. INTRODUCTION

Time-of-Flight (ToF) depth cameras have been used for various 3D tasks such as 3D object modeling [1] and tracking [2] because they provide real-time 3D information at each pixel.

The ToF depth camera is composed of a light emitter and a light detector. The emitted optical signal is reflected by objects within the illuminated scene and travels back to the camera. The travel time, or the phase delay of the incoming signal, is computed by correlating the emitted signal and the incoming signal [3], based on the assumption that every illuminated object is within a certain maximum range. The range measurement $R$ at each pixel is then given by:

$$R = R_{\text{max}} \frac{\phi}{2\pi}.$$  

Here, $R_{\text{max}}$ is the so-called non-ambiguous distance range, which is the maximum range inversely proportional to the modulation frequency. $\phi \in [0, 2\pi)$ is the phase delay of the incoming signal.

Because the signals are continuously emitted and incoming, $\phi$ is inherently indistinguishable from $\phi + 2\pi n$, where $n$ is the number of mods. For objects out of the non-ambiguous distance range, $n$ should be positive, but $n$ is unknown and not given by the camera. This ambiguity causes $R$ to be indistinguishable from $R + nR_{\text{max}}$. For example, Fig. 1 shows the cases where $n$ is 0, 1, and 2, and the corresponding range values are also illustrated.

This paper aims at extending the range use of a ToF depth camera beyond the non-ambiguous distance range by estimating $n$ at each pixel, or equivalently, by range unfolding. To our best knowledge, we propose the first attempt for range unfolding, given a single range image and its associated amplitude image, the intensity of the incoming light, acquired by a ToF depth camera. Considering $n$ as a label of each pixel, range unfolding is a labeling problem, which can be solved in the Markov random field (MRF) framework [4]. In this framework, we define an energy function based on an assumption on the statistics of the amplitude and range values, and another assumption on the continuity in the range values of neighboring pixels. The actual range estimates are then recovered by finding the labels that minimize the energy.

The previous methods [5, 6] for range unfolding use two or more range images of the same scene acquired under different modulation frequencies. While successful in many static scenes, these methods remain difficult to dynamic environments where concurrent acquisition of multiple images of the same scene is hardly possible. On the other hand, our method...
is readily applicable for such situations due to the use of a single range image.

The rest of this paper is organized as follows. The next section presents our range unfolding method. Section 3 shows the validity of our approach through experiments. Finally, conclusions and future work are provided in Section 4.

2. RANGE UNFOLDING VIA MARKOV RANDOM FIELD MODELING

Range unfolding from a single range image is inherently an ill-posed problem, without any additional prior knowledge or assumptions. Fortunately, useful assumptions can be made in indoor environments. First, we assume that the scale of the environment is roughly known, so that $M$, the maximum value of $n$, can be fixed. Under this assumption, range unfolding can be considered as a labeling problem assigning a label $n \in \{0, \ldots, M\}$ to each pixel.

The second assumption is that the corrected amplitude values of pixels with $n = 0$ are higher than those of pixels with $n > 0$, where the corrected amplitude value is the product of the amplitude value and the squared range value \cite{7}. This correction amplifies the amplitude value of a pixel with correct range measurement. Fig. 2(c) shows a corrected amplitude image in which this assumption holds.

The third assumption is that scene surfaces are continuous, so that the range values are also continuous except at the folding boundaries where the range values exhibit large differences near $R_{\text{max}}$. Finally, the fourth assumption is that these boundaries are visible in the range image. Fig. 2(a) shows a range image in which these assumptions hold.

The second assumption describes the observation at each pixel, and the third and fourth assumptions on range continuity explain the observation between two neighboring pixels minimally. Thus, we can construct a Markov random field \cite{4} with energy function $E$:

$$E(n) = \sum_{i \in S} D_i(n_i) + \sum_{i \in S} \sum_{j \in N_i} V(n_i, n_j),$$

where $n$ is the set of labels, $i$ and $j$ denote pixels, $S$ is the set of pixels, and $N_i$ is the pixels in the eight-neighborhood of pixel $i$. $D_i(n_i)$ is the cost of assigning label $n_i$ to pixel $i$, and is referred to as the data cost. $V(n_i, n_j)$ is the cost of assigning labels $n_i$ and $n_j$ to two neighboring pixels, and is referred to as the discontinuity cost.

Now, our goal in this section is to define each cost appropriately so that the range measurements can be unfolded by finding the optimal $n^*$ that minimizes $E$. In the following subsection we define $D_i(n_i)$ based on the second assumption, and in Subsection 2.2, we define $V(n_i, n_j)$ based on the third and fourth assumptions.

2.1. Data cost

We adaptively define our data cost according to the corrected amplitude value at each pixel. For this purpose, we first classify the pixels into two kinds: pixels $H$ with high corrected amplitude values and pixels $L$ with low corrected amplitude values. According to the second assumption, $n$ is likely to be zero at pixels $H$, while a positive integer at pixels $L$.

Due to the fact that the corrected amplitude value can be affected by high or low reflectance of objects, we classify the pixels with extremely high and low amplitude values as a separate set $U$, apart from $L$ and $H$. Fig. 3(a) displays the pix-
els classified as $U$, which have been projected from the light sources.

The threshold value that separates $H$ from $L$ should be adaptively determined because the distribution of corrected amplitude values is different from image to image. We fit a Gaussian mixture model with two components to the histogram of the corrected amplitude values of all the pixels except $U$:

$$p(I_i | \Theta) = \sum_{k=1}^{2} \alpha_k p_k(I_i | \theta_k),$$  

where $I_i$ denotes the corrected amplitude value of pixel $i$, $\alpha_k$ is a nonnegative value such that $\sum \alpha_k = 1$. $p_k(I_i | \theta_k)$ is a Gaussian distribution with parameters $\theta_k$, consisting of its mean and variance. The parameters $\Theta = \{\alpha_k, \theta_k | k = 1, 2\}$ are estimated by the Expectation Maximization (EM) algorithm [8]. Assigning $H$ to the component distribution with higher mean value, and $L$ to the other, we then classify a pixel as $H$ if $P(H | I_i) = \frac{\alpha_H p(H | \theta_H)}{p(H | \Theta)} > 0.5$. Fig. 3(b) shows the classification result, where small misclassification occurs at the regions of the self-emitting objects such as the lights on the ceiling.

To obtain more accurate results, we use a technique based on a graph-cut based image segmentation algorithm [9]. Fig. 3(c) shows the better classification result. In addition, we add the pixels in the vicinity of the label boundaries to $U$ (refer to Fig. 3(d), for example), to prevent possible error at the boundaries from propagating to the next step.

Finally, we define $D_i(n_i)$ adaptively to the above classification result. Because our second assumption suggests that pixels with $n = 0$ are likely to belong to $H$ with large $P(H | I_i)$, we define $D_i(n_i)$ for $i \in H$ as:

$$D_i(n_i) = \begin{cases} 1 - P(H | I_i) & \text{if } n_i = 0, \\ 1 & \text{otherwise}. \end{cases}$$  

On the other hand, assuming equal probability for each $n_i > 0$, we define $D_i(n_i)$ for $i \in L$ as:

$$D_i(n_i) = \begin{cases} 1 & \text{if } n_i = 0, \\ 1 - \frac{p(L | I_i)}{M} & \text{otherwise}. \end{cases}$$  

For all the pixels in $U$, we set $D_i(n_i)$ to 1 regardless of $n_i$ because we consider these pixels do not follow the second assumption.

2.2. Discontinuity cost

Under the third assumption, range values should be continuous almost everywhere in the range image. Folding boundaries, however, exhibit large range differences if they are visible in the range image, as shown in Fig. 2(a). Based on this observation, we can find the label difference $c^*_{ij}$ minimizing the range discontinuity between every neighboring pixels $i$ and $j$:

$$c^*_{ij} = \arg\min_{c \in \{-1, 0, 1\}} |R_i - R_j + R_{\text{max}}c|,$$  

where $R_i$ and $R_j$ denote the range measurement at pixels $i$ and $j$, respectively.

In order for the range values to be as continuous as possible, these optimal label differences should be preserved at every neighboring pixel pair; however, the differences may not be consistent across pixel pairs because of noise and occlusion. Thus, the differences should be globally optimized under our MRF framework by defining $V(n_i, n_j)$ in the manner of penalizing $i$ and $j$ if $n_i - n_j \neq c^*_{ij}$:

$$V(n_i, n_j) = \begin{cases} \gamma \exp(-\beta R_{ij}^2) & \text{if } R_{ij}^* < \tau, \\ 0 & \text{otherwise}, \end{cases}$$  

where $R_{ij}^* = |R_i - R_j + R_{\text{max}}c^*_{ij}|$, and $\gamma$ and $\beta$ are positive constants. $\tau$ is a threshold value for excluding originally discontinuous pixel pairs.

We can find $\mathbf{n}^*$ that minimizes $E$ using the belief propagation algorithm [10]. We then recover the actual range estimates using $\mathbf{n}^*$ as shown in Fig. 2(d).

3. EXPERIMENTS

We have tested our method using 25 sample indoor images, where $M = 1$ for 16 images and $M = 2$ for the remains, with $R_{\text{max}} = 5m$. For evaluation, the ground-truth of each range image is estimated using multiple modulation frequencies [5].

The success rate is computed as the ratio of the pixels with correct label $n$ in each range image. The average success rate is 92.4%, while 57.7% of the range measurements are originally correct on average, without unfolding. This result shows that our method extends the range use of a ToF depth camera beyond $R_{\text{max}}$ at least twice via elimination of the ambiguity from unknown $n$.

Fig. 4 shows several input images and their results involved in our experiments. In Fig. 4(a) the result is 99.7% correct because our assumptions strongly hold. Fig. 4(b, c) show a case in which our assumptions weakly hold: the change in range measurements is not large at the folding boundary in the left part of Fig. 4(b) due to occlusion. Fig. 4(d) shows a case in which not only the second assumption fails but also the graph-cut algorithm over-smoothes the EM result. For this reason, originally correct range measurements are partially false-unfolded, resulting in the worst success rate. The regions of the hair and monitor, which have not been classified as $U$ in spite of their low amplitude values, affect the body region to be classified as $L$. This kind of effect can be also observed in the lower left part of Fig. 4(c).

4. CONCLUSIONS AND FUTURE WORK

We have proposed a method for range unfolding, which is, to our knowledge, the first method using a single range image and its associated amplitude image. Our approach recasts range unfolding as an optimization problem in the Markov

4191
Fig. 4. Input images and their results involved in our experiments. The ratio of pixels with originally correct range measurements is displayed below each range image, while the success rate is displayed below each unfolding result.

random field framework via modeling the costs involving one pixel or two neighboring pixels. Through experiments, we have shown that our method can restore the actual range estimates from ambiguous range measurements from ToF depth cameras, consequently extending the range use at least twice.

Our assumptions may not cover all the variety of real scenes. To generalize the assumptions, we are currently working for a method using an image sequence acquired in dynamic situations.

5. REFERENCES


