

Chapter 2

Disambiguation of Time-of-Flight Data

Abstract The maximum range of a time-of-flight camera is limited by the periodicity of the measured signal. Beyond a certain range, which is determined by the signal frequency, the measurements are confounded by phase wrapping. This effect is demonstrated in real examples. Several phase-unwrapping methods, which can be used to extend the range of time-of-flight cameras, are discussed. Simple methods can be based on the measured amplitude of the reflected signal, which is itself related to the depth of objects in the scene. More sophisticated unwrapping methods are based on zero-curl constraints, which enforce spatial consistency on the phase measurements. Alternatively, if more than one depth camera is used, then the data can be unwrapped by enforcing consistency among different views of the same scene point. The relative merits and shortcomings of these methods are evaluated, and the prospects for hardware-based approaches, involving frequency modulation are discussed.

Keywords Time-of-Flight principle · Depth ambiguity · Phase unwrapping · Multiple depth cameras

2.1 Introduction

Time-of-Flight cameras emit modulated infrared light and detect its reflection from the illuminated scene points. According to the ToF principle described in Chap. 1, the detected signal is gated and integrated using internal reference signals, to form the tangent of the phase ϕ of the detected signal. Since the tangent of ϕ is a periodic function with a period of 2π , the value $\phi + 2n\pi$ gives exactly the same tangent value for any nonnegative integer n .

Commercially available ToF cameras compute ϕ on the assumption that ϕ is within the range of $[0, 2\pi)$. For this reason, each modulation frequency f has its maximum range d_{\max} corresponding to 2π , encoded without ambiguity:

$$d_{\max} = \frac{c}{2f}, \quad (2.1)$$

where c is the speed of light. For any scene points farther than d_{\max} , the measured distance d is much shorter than its actual distance $d + nd_{\max}$. This phenomenon is called *phase wrapping*, and estimating the unknown number of wrappings n is called *phase unwrapping*.

For example, the Mesa SR4000 [16] camera records a 3D point \mathbf{X}_p at each pixel p , where the measured distance d_p equals $\|\mathbf{X}_p\|$. In this case, the unwrapped 3D point $\mathbf{X}_p(n_p)$ with number of wrappings n_p can be written as

$$\mathbf{X}_p(n_p) = \frac{d_p + n_p d_{\max}}{d_p} \mathbf{X}_p. \quad (2.2)$$

Figure 2.1a shows a typical depth map acquired by the SR4000 [16], and Fig. 2.1b shows its unwrapped depth map. As shown in Fig. 2.1e, phase unwrapping is crucial for recovering large-scale scene structure.

To increase the usable range of ToF cameras, it is also possible to extend the maximum range d_{\max} by decreasing the modulation frequency f . In this case, the integration time should also be extended, to acquire a high quality depth map, since the depth noise is inversely proportional to f . With extended integration time, moving objects are more likely to result in motion artifacts. In addition, we do not know at which modulation frequency phase wrapping does not occur, without exact knowledge regarding the scale of the scene.

If we can accurately unwrap a depth map acquired at a high modulation frequency, then the unwrapped depth map will suffer less from noise than a depth map acquired at a lower modulation frequency, integrated for the same time. Also, if a phase-unwrapping method does not require exact knowledge on the scale of the scene, then the method will be applicable in more large-scale environments.

There exist a number of phase-unwrapping methods [4–8, 14, 17, 21] that have been developed for ToF cameras. According to the number of input depth maps, the methods are categorized into two groups: those using a single depth map [5, 7, 14, 17, 21] and those using multiple depth maps [4, 6, 8, 20]. The following subsections introduce their principles, advantages and limitations.

2.2 Phase Unwrapping from a Single Depth Map

ToF cameras such as the SR4000 [16] provide an amplitude image along with its corresponding depth map. The amplitude image is encoded with the strength of the detected signal, which is inversely proportional to the squared distance. To obtain *corrected amplitude* A' [19], which is proportional to the reflectivity of a scene surface with respect to the infrared light, we can multiply amplitude A and its corresponding squared distance d^2 :

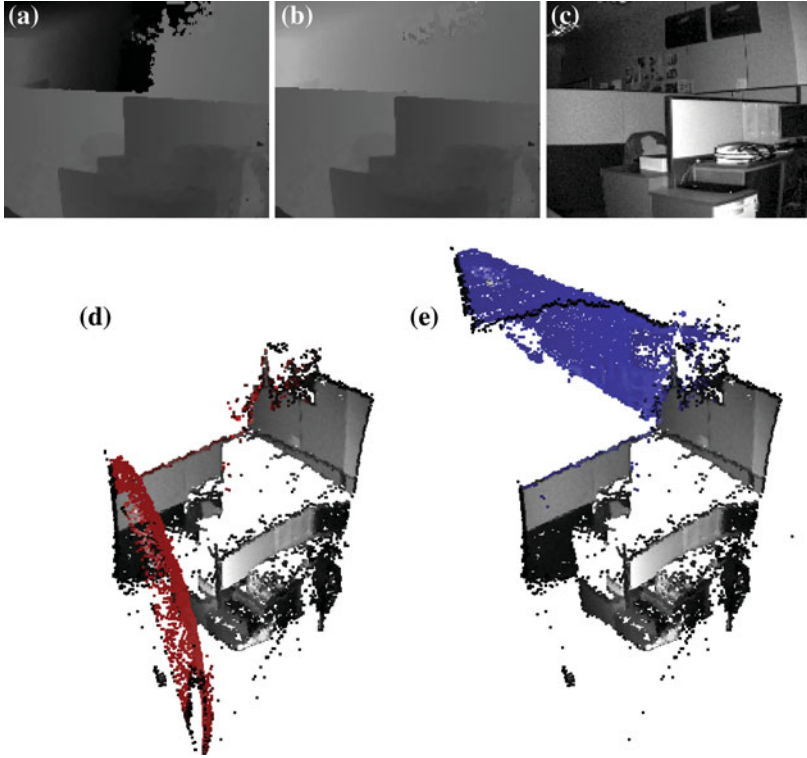


Fig. 2.1 Structure recovery through phase unwrapping. **a** Wrapped ToF depth map. **b** Unwrapped depth map corresponding to (a). Only the distance values are displayed in (a) and (b), to aid visibility. The intensity is proportional to the distance. **c** Amplitude image associated with (a). **d** and **e** display the 3D points corresponding to (a) and (b), respectively. **d** The wrapped points are displayed in *red*. **e** Their unwrapped points are displayed in *blue*. The remaining points are textured using the original amplitude image (c)

$$A' = Ad^2. \quad (2.3)$$

Figure 2.2 shows an example of amplitude correction. It can be observed from Fig. 2.2c that the corrected amplitude is low in the wrapped region. Based on the assumption that the reflectivity is constant over the scene, the corrected amplitude values can play an important role in detecting wrapped regions [5, 17, 21].

Poppinga and Birk [21] use the following inequality for testing if the depth of pixel p has been wrapped:

$$A'_p \leq A_p^{\text{ref}} T, \quad (2.4)$$

where T is a manually chosen threshold, and A_p^{ref} is the reference amplitude of pixel p when viewing a white wall at 1 m, approximated by

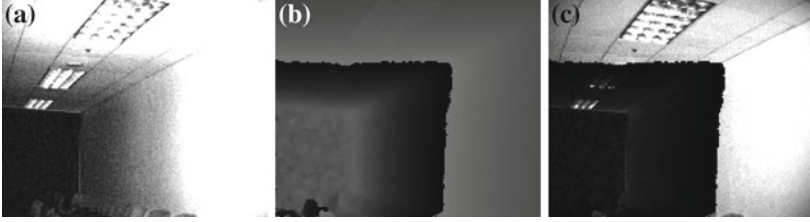


Fig. 2.2 Amplitude correction example. **a** Amplitude image. **b** TOF depth map. **c** Corrected amplitude image. The intensity in (b) is proportional to the distance. The lower left part of (b) has been wrapped. Images courtesy of Choi et al. [5]

$$A_p^{\text{ref}} = B - ((x_p - c_x)^2 + (y_p - c_y)^2), \quad (2.5)$$

where B is a constant. The image coordinates of p are (x_p, y_p) , and (c_x, c_y) is approximately the image center, which is usually better illuminated than the periphery. A_p^{ref} compensates this effect by decreasing $A_p^{\text{ref}} T$ if pixel p is in the periphery.

After the detection of wrapped pixels, it is possible to directly obtain an unwrapped depth map by setting the number of wrappings of the wrapped pixels to one on the assumption that the maximum number of wrappings is 1.

The assumption on the constant reflectivity tends to be broken when the scene is composed of different objects with varying reflectivity. This assumption cannot be fully relaxed without detailed knowledge of scene reflectivity, which is hard to obtain in practice. To robustly handle varying reflectivity, it is possible to adaptively set the threshold for each image and to enforce spatial smoothness on the detection results.

Choi et al. [5] model the distribution of corrected amplitude values in an image using a mixture of Gaussians with two components, and apply expectation maximization [1] to learn the model:

$$p(A'_p) = \alpha_H p(A'_p | \mu_H, \sigma_H^2) + \alpha_L p(A'_p | \mu_L, \sigma_L^2), \quad (2.6)$$

where $p(A'_p | \mu, \sigma^2)$ denotes a Gaussian distribution with mean μ and variance σ^2 , and α is the coefficient for each distribution. The components $p(A'_p | \mu_H, \sigma_H^2)$ and $p(A'_p | \mu_L, \sigma_L^2)$ describe the distributions of high and low corrected amplitude values, respectively. Similarly, the subscripts H and L denote labels *high* and *low*, respectively. Using the learned distribution, it is possible to write a probabilistic version of Eq. (2.4) as

$$P(H | A'_p) < 0.5, \quad (2.7)$$

where $P(H | A'_p) = \alpha_H p(A'_p | \mu_H, \sigma_H^2) / p(A'_p)$.

To enforce spatial smoothness on the detection results, Choi et al. [5] use a segmentation method [22] based on Markov random fields (MRFs). The method finds the binary labels $n \in \{H, L\}$ or $\{0, 1\}$ that minimize the following energy:

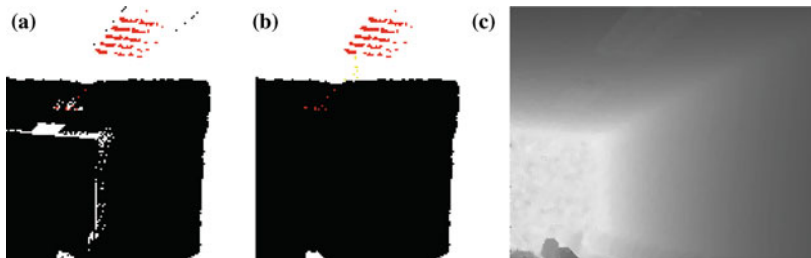


Fig. 2.3 Detection of wrapped regions. **a** Result obtained by expectation maximization. **b** Result obtained by MRF optimization. The pixels with labels L and H are colored in *black* and *white*, respectively. The *red* pixels are those with extremely high or low amplitude values, which are not processed during the classification. **c** Unwrapped depth map corresponding to Fig. 2.2(b). The intensity is proportional to the distance. Images courtesy of Choi et al. [5]

$$E = \sum_p D_p(n_p) + \sum_{(p,q)} V(n_p, n_q), \quad (2.8)$$

where $D_p(n_p)$ is a data cost that is defined as $1 - P(n_p|A'_p)$, and $V(n_p, n_q)$ is a discontinuity cost that penalizes a pair of adjacent pixels p and q if their labels n_p and n_q are different. $V(n_p, n_q)$ is defined in a manner of increasing the penalty if a pair of adjacent pixels have similar corrected amplitude values:

$$V(n_p, n_q) = \lambda \exp(-\beta(A'_p - A'_q)^2) \delta(n_p \neq n_q), \quad (2.9)$$

where λ and β are constants, which are either manually chosen or adaptively determined. $\delta(x)$ is a function that evaluates to 1 if its argument is true and evaluates to zero otherwise.

Figure 2.3 shows the classification results obtained by Choi et al. [5]. Because of varying reflectivity of the scene, the result in Fig. 2.3a exhibits misclassified pixels in the lower left part. The misclassification is reduced by applying the MRF optimization as shown in Fig. 2.3b. Figure 2.3c shows the unwrapped depth map obtained by Choi et al. [5], corresponding to Fig. 2.2b.

McClure et al. [17] also use a segmentation-based approach, in which the depth map is segmented into regions by applying the watershed transform [18]. In their method, wrapped regions are detected by checking the average corrected amplitude of each region.

On the other hand, depth values tend to be highly discontinuous across the wrapping boundaries, where there are transitions in the number of wrappings. For example, the depth maps in Figs. 2.1a, 2.2b shows such discontinuities. On the assumption that the illuminated surface is smooth, the depth difference between adjacent pixels should be small. If the difference between measured distances is greater than $0.5d_{\max}$ for any adjacent pixels, say $d_p - d_q > 0.5d_{\max}$, we can set the number of relative wrappings, or, briefly, the shift $n_q - n_p$ to 1 so that the unwrapped difference will satisfy $-0.5d_{\max} \leq d_p - d_q - (n_q - n_p)d_{\max} < 0$, minimizing the discontinuity.



Fig. 2.4 One-dimensional phase-unwrapping example. **a** Measured phase image. **b** Unwrapped phase image where the phase difference between p and q is now *less* than 0.5. In (a) and (b), all the phase values have been divided by 2π . For example, the displayed value 0.1 corresponds to 0.2π

Fig. 2.5 Two-dimensional phase-unwrapping example. **a** Measured phase image. **(b–d)** Sequentially unwrapped phase images where the phase difference across the *red dotted line* has been minimized. From **a** to **d**, all the phase values have been divided by 2π . For example, the displayed value 0.1 corresponds to 0.2π

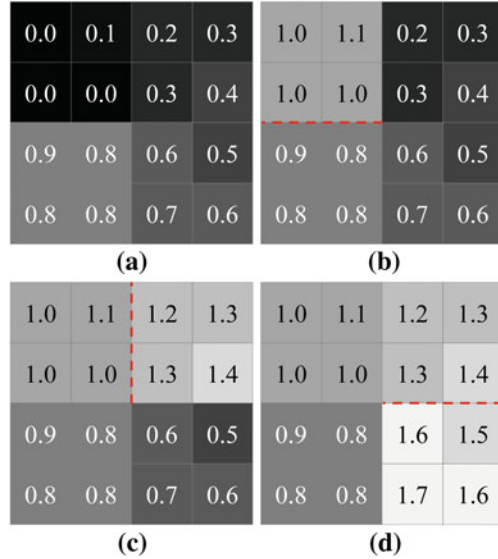


Figure 2.4 shows a one-dimensional phase-unwrapping example. In Fig. 2.4a, the phase difference between pixels p and q is greater than 0.5 (or π). The shifts that minimize the difference between adjacent pixels are 1 (or, $n_q - n_p = 1$) for p and q , and 0 for the other pairs of adjacent pixels. On the assumption that n_p equals 0, we can integrate the shifts from left to right to obtain the unwrapped phase image in Fig. 2.4b.

Figure 2.5 shows a two-dimensional phase-unwrapping example. From Fig. 2.5a to d, the phase values are unwrapped in a manner of minimizing the phase difference across the red dotted line. In this two-dimensional case, the phase differences greater than 0.5 never vanish, and the red dotted line cycles around the image center infinitely. This is because of the local phase error that causes the violation of the *zero-curl* constraint [9, 12].

Figure 2.6 illustrates the zero-curl constraint. Given four neighboring pixel locations (x, y) , $(x + 1, y)$, $(x, y + 1)$, and $(x + 1, y + 1)$, let $a(x, y)$ and $b(x, y)$ denote the shifts $n(x + 1, y) - n(x, y)$ and $n(x, y + 1) - n(x, y)$, respectively, where $n(x, y)$ denotes the number of wrappings at (x, y) . Then, the shift $n(x + 1, y + 1) - n(x, y)$ can be calculated in two different ways: either $a(x, y) + b(x + 1, y)$ or $b(x, y) + a(x, y + 1)$

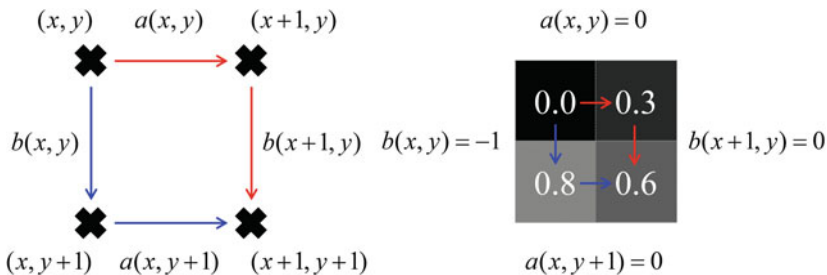


Fig. 2.6 Zero-curl constraint: $a(x, y) + b(x + 1, y) = b(x, y) + a(x, y + 1)$. **a** The number of relative wrappings between $(x + 1, y + 1)$ and (x, y) should be consistent regardless of its integrating paths. For example, two different paths (*red* and *blue*) are shown. **b** shows an example in which the constraint is not satisfied. The *four* pixels correspond to the *four* pixels in the middle of Fig. 2.5a

following one of the two different paths shown in Fig. 2.6a. For any phase-unwrapping results to be consistent, the two values should be the same, satisfying the following equality:

$$a(x, y) + b(x + 1, y) = b(x, y) + a(x, y + 1). \quad (2.10)$$

Because of noise or discontinuities in the scene, the zero-curl constraint may not be satisfied locally, and the local error is propagated to the entire image during the integration. There exist classical phase-unwrapping methods [9, 12] applied in magnetic resonance imaging [15] and interferometric synthetic aperture radar (SAR) [13], which rely on detecting [12] or fixing [9] broken zero-curl constraints. Indeed, these classical methods [9, 12] have been applied to phase unwrapping for ToF cameras [7, 14].

2.2.1 Deterministic Methods

Goldstein et al. [12] assume that the shift is either 1 or -1 between adjacent pixels if their phase difference is greater than π , and assume that it is 0 otherwise. They detect cycles of four neighboring pixels, referred to as *plus and minus residues*, which do not satisfy the zero-curl constraint.

If any integration path encloses an unequal number of plus and minus residue, the integrated phase values on the path suffer from global errors. In contrast, if any integration path encloses an equal number of plus and minus residues, the global error is balanced out. To prevent global errors from being generated, Goldstein et al. [12] connect nearby plus and minus residues with *cuts*, which interdict the integration paths, such that no net residues can be encircled.

After constructing the cuts, the integration starts from a pixel p , and each neighboring pixel q is unwrapped relatively to p in a greedy and sequential manner if q has not been unwrapped and if p and q are on the same side of the cuts.

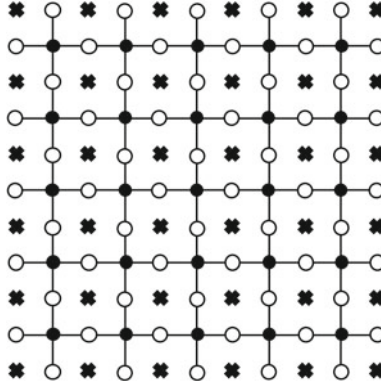


Fig. 2.7 Graphical model that describes the zero-curl constraints (*black discs*) between neighboring shift variables (*white discs*). 3-element probability vectors (μ 's) on the shifts between adjacent nodes ($-1, 0, \text{ or } 1$) are propagated across the network. The x marks denote pixels [9]

2.2.2 Probabilistic Methods

Frey et al. [9] propose a very loopy belief propagation method for estimating the shift that satisfies the zero-curl constraints. Let the set of shifts, and a measured phase image, be denoted by

$$S = \left\{ a(x, y), b(x, y) : x = 1, \dots, N-1; y = 1, \dots, M-1 \right\}$$

and

$$\Phi = \left\{ \phi(x, y) : 0 \leq \phi(x, y) < 1, x = 1, \dots, N; y = 1, \dots, M \right\},$$

respectively, where the phase values have been divided by 2π . The estimation is then recast as finding the solution that maximizes the following joint distribution:

$$p(S, \Phi) \propto \prod_{x=1}^{N-1} \prod_{y=1}^{M-1} \delta(a(x, y) + b(x+1, y) - a(x, y+1) - b(x, y))$$

$$\times \prod_{x=1}^{N-1} \prod_{y=1}^M e^{-\left(\phi(x+1, y) - \phi(x, y) + a(x, y)\right)^2 / 2\sigma^2}$$

$$\times \prod_{x=1}^N \prod_{y=1}^{M-1} e^{-\left(\phi(x, y+1) - \phi(x, y) + b(x, y)\right)^2 / 2\sigma^2}$$

where $\delta(x)$ evaluates to 1 if $x = 0$ and to 0 otherwise. The variance σ^2 is estimated directly from the wrapped phase image [9].

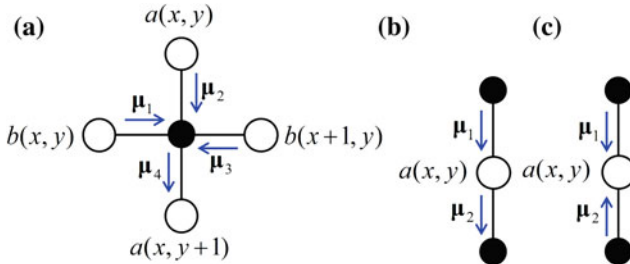


Fig. 2.8 **a** Constraint-to-shift vectors are computed from incoming shift-to-constraint vectors. **b** Shift-to-constraint vectors are computed from incoming constraint-to-shift vectors. **c** Estimates of the marginal probabilities of the shifts given the data are computed by combining incoming constraint-to-shift vectors [9]

Frey et al. [9] construct a graphical model describing the factorization of $p(S, \Phi)$, as shown in Fig. 2.7. In the graph, each shift node (white disc) is located between two pixels, and corresponds to either an x -directional shift (a 's) or a y -directional shift (b 's). Each constraint node (black disc) corresponds to a zero-curl constraint, and is connected to its four neighboring shift nodes. Every node passes a message to its neighboring node, and each message is a 3-vector denoted by μ , whose elements correspond to the allowed values of shifts, -1 , 0 , and 1 . Each element of μ can be considered as a probability distribution over the three possible values [9].

Figure 2.8a illustrates the computation of a message μ_4 from a constraint node to one of its neighboring shift nodes. The constraint node receives messages μ_1 , μ_2 , and μ_3 from the rest of its neighboring shift nodes, and filters out the joint message elements that do not satisfy the zero-curl constraint:

$$\mu_{4i} = \sum_{j=-1}^1 \sum_{k=-1}^1 \sum_{l=-1}^1 \delta(k+l-i-j) \mu_{1j} \mu_{2k} \mu_{3l}, \quad (2.11)$$

where μ_{4i} denotes the element of μ_4 , corresponding to shift value $i \in \{-1, 0, 1\}$.

Figure 2.8b illustrates the computation of a message μ_2 from a shift node to one of its neighboring constraint nodes. Among the elements of the message μ_1 from the other neighboring constraint node, the element, which is consistent with the measured shift $\phi(x, y) - \phi(x+1, y)$, is amplified:

$$\mu_{2i} = \mu_{1i} \exp\left(-(\phi(x+1, y) - \phi(x, y) + i)^2 / 2\sigma^2\right). \quad (2.12)$$

After the messages converge (or, after a fixed number of iterations), an estimate of the marginal probability of a shift is computed by using the messages passed into its corresponding shift node, as illustrated in Fig. 2.8c:

$$\hat{P}(a(x, y) = i | \Phi) = \frac{\mu_{1i} \mu_{2i}}{\sum_j \mu_{1j} \mu_{2j}}. \quad (2.13)$$

Given the estimates of the marginal probabilities, the most probable value of each shift node is selected. If some zero-curl constraints remain violated, a robust integration technique, such as least-squares integration [10] should be used [9].

2.2.3 Discussion

The aforementioned phase-unwrapping methods using a single depth map [5, 7, 14, 17, 21] have an advantage that the acquisition time is not extended, keeping the motion artifacts at a minimum. The methods, however, rely on strong assumptions that are fragile in real world situations. For example, the reflectivity of the scene surface may vary in a wide range. In this case, it is hard to detect wrapped regions based on the corrected amplitude values. In addition, the scene may be discontinuous if it contains multiple objects that occlude one another. In this case, the wrapping boundaries tend to coincide with object boundaries, and it is often hard to observe large depth discontinuities across the boundaries, which play an important role in determining the number of relative wrappings.

The assumptions can be relaxed by using multiple depth maps at a possible extension of acquisition time. The next subsection introduces phase-unwrapping methods using multiple depth maps.

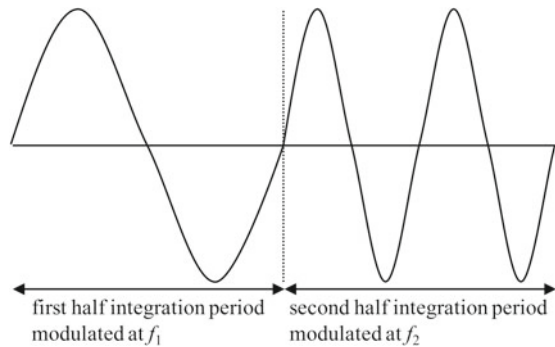
2.3 Phase Unwrapping from Multiple Depth Maps

Suppose that a pair of depth maps M_1 and M_2 of a static scene are given, which have been taken at different modulation frequencies f_1 and f_2 from the same viewpoint. In this case, pixel p in M_1 corresponds to pixel p in M_2 , since the corresponding region of the scene is projected onto the same location of M_1 and M_2 . Thus, the unwrapped distances at those corresponding pixels should be consistent within the noise level.

Without prior knowledge, the noise in the unwrapped distance can be assumed to follow a zero-mean distribution. Under this assumption, the maximum likelihood estimates of the numbers of wrappings at the corresponding pixels should minimize the difference between their unwrapped distances. Let m_p and n_p be the numbers of wrappings at pixel p in M_1 and M_2 , respectively. Then, we can choose m_p and n_p that minimize $g(m_p, n_p)$ such that

$$g(m_p, n_p) = |d_p(f_1) + m_p d_{\max}(f_1) - d_p(f_2) - n_p d_{\max}(f_2)|, \quad (2.14)$$

Fig. 2.9 Frequency modulation within an integration period. The first half is modulated at f_1 , and the other half is modulated at f_2



where $d_p(f_1)$ and $d_p(f_2)$ denote the measured distances at pixel p in M_1 and M_2 respectively, and $d_{\max}(f)$ denotes the maximum range of f .

The depth consistency constraint has been mentioned by Göktürk et al. [11] and used by Falie and Buzuloiu [8] for phase unwrapping of ToF cameras. The illuminating power of ToF cameras is, however, limited due to the eye-safety problem, and the reflectivity of the scene may be very low. In this situation, the amount of noise may be too large for accurate numbers of wrappings to minimize $g(m_p, n_p)$. For robust estimation against noise, Droeschel et al. [6] incorporate the depth consistency constraint into their earlier work [7] for a single depth map, using an auxiliary depth map of a different modulation frequency.

If we acquire a pair of depth maps of a dynamic scene sequentially and independently, the pixels at the same location may not correspond to each other. To deal with such dynamic situations, several approaches [4, 20] acquire a pair of depth maps simultaneously. These can be divided into single-camera and multicamera methods, as described below.

2.3.1 Single-Camera Methods

For obtaining a pair of depth maps sequentially, four samples of integrated electric charge are required per each integration period, resulting in eight samples within a pair of two different integration periods. Payne et al. [20] propose a special hardware system that enables simultaneous acquisition of a pair of depth maps at different frequencies by dividing the integration period into two, switching between frequencies f_1 and f_2 , as shown in Fig. 2.9.

Payne et al. [20] also shows that it is possible to obtain a pair of depth maps with only five or six samples within a combined integration period, using their system. By using fewer samples, the total readout time is reduced and the integration period for each sample can be extended, resulting in an improved signal-to-noise ratio.

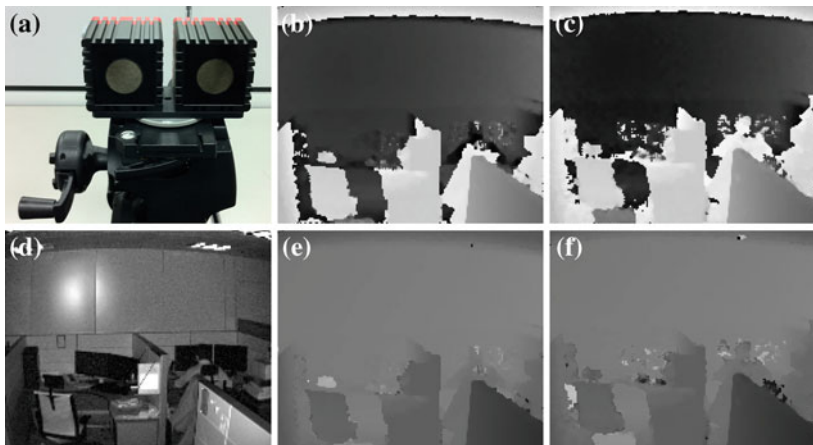


Fig. 2.10 **a** Stereo ToF camera system. **(b, c)** Depth maps acquired by the system. **d** Amplitude image corresponding to (b). **(e, f)** Unwrapped depth maps, corresponding to (b) and (c), respectively. The intensity in (b, c, e, f) is proportional to the depth. The maximum intensity (255) in (b, c) and (e, f) correspond to 5.2 and 15.6 m, respectively. Images courtesy of Choi and Lee [4]

2.3.2 Multicamera Methods

Choi and Lee [4] use a *pair* of commercially available ToF cameras to simultaneously acquire a pair of depth maps from different viewpoints. The two cameras C_1 and C_2 are fixed to each other, and the mapping of a 3D point \mathbf{X} from C_1 to its corresponding point \mathbf{X}' from C_2 is given by (\mathbf{R}, \mathbf{T}) , where \mathbf{R} is a 3×3 rotation matrix, and \mathbf{T} is a 3×1 translation vector. In [4], the extrinsic parameters \mathbf{R} and \mathbf{T} are assumed to have been estimated. Figure 2.10a shows the stereo ToF camera system.

Denoting by M_1 and M_2 a pair of depth maps acquired by the system, a pixel p in M_1 and its corresponding pixel q in M_2 should satisfy:

$$\mathbf{X}'_q(n_q) = \mathbf{R}\mathbf{X}_p(m_p) + \mathbf{T}, \quad (2.15)$$

where $\mathbf{X}_p(m_p)$ and $\mathbf{X}'_q(n_q)$ denote the unwrapped 3D points of p and q with their numbers of wrappings m_p and n_q , respectively.

Based on the relation in Eq.(2.15), Choi and Lee [4] generalize the depth consistency constraint in Eq.(2.14) for a single camera to those for the stereo camera system:

$$D_p(m_p) = \min_{n_{q^*} \in \{0, \dots, N\}} \left(\left\| \mathbf{X}'_{q^*}(n_{q^*}) - \mathbf{R}\mathbf{X}_p(m_p) - \mathbf{T} \right\| \right), \quad (2.16)$$

$$D_q(n_q) = \min_{m_{p^*} \in \{0, \dots, N\}} \left(\left\| \mathbf{X}_{p^*}(m_{p^*}) - \mathbf{R}^T(\mathbf{X}'_q(n_q) - \mathbf{T}) \right\| \right),$$

Table 2.1 Summary of phase-unwrapping methods

Methods	# Depth maps	Cues	Approach	Maximum range
Poppinga and Birk [21]	1	CA ^a	Thresholding	$2d_{\max}$
Choi et al. [5]	1	CA, DD ^b	Segmentation, MRF	$(N^d + 1)d_{\max}$
McClure et al. [17]	1	CA	Segmentation, thresholding	$2d_{\max}$
Jutzi [14]	1	DD	Branch cuts, integration	∞
Droeschel et al. [7]	1	DD	MRF, integration	∞
Droeschel et al. [6]	2 (Multi-freq.)	DD, DC ^c	MRF, integration	∞
Payne et al. [20]	2 (Multi-freq.)	DC	Hardware	$\frac{c}{2 f_1 - f_2 }$
Choi and Lee [4]	2 (Stereo)	DC	Stereo ToF, MRF	$(N + 1)d_{\max}$

^a Corrected amplitude. ^b Depth discontinuity. ^c Depth consistency. ^d The maximum number of wrappings determined by the user.

where pixels q^* and p^* are the projections of $\mathbf{R}\mathbf{X}_p(m_p) + \mathbf{T}$ and $\mathbf{R}^T(\mathbf{X}'_q(n_q) - \mathbf{T})$ onto M_2 and M_1 , respectively. The integer N is the maximum number of wrappings, determined by approximate knowledge on the scale of the scene.

To robustly handle with noise and occlusion, Choi and Lee [4] minimize the following MRF energy functions E_1 and E_2 , instead of independently minimizing $D_p(m_p)$ and $D_q(m_q)$ at each pixel:

$$E_1 = \sum_{p \in M_1} \hat{D}_p(m_p) + \sum_{(p,u)} V(m_p, m_u), \quad (2.17)$$

$$E_2 = \sum_{q \in M_2} \hat{D}_q(n_q) + \sum_{(q,v)} V(n_q, n_v),$$

where $\hat{D}_p(m_p)$ and $\hat{D}_q(n_q)$ are the data cost of assigning m_p and n_q to pixels p and q , respectively. Functions $V(m_p, m_u)$ and $V(n_q, n_v)$ determine the discontinuity cost of assigning (m_p, m_u) and (n_q, n_v) to pairs of adjacent pixels (p, u) and (q, v) , respectively.

The data costs $\hat{D}_p(m_p)$ and $\hat{D}_q(n_q)$ are defined by truncating $D_p(m_p)$ and $D_q(n_q)$ to prevent their values from becoming too large, due to noise and occlusion:

$$\hat{D}_p(m_p) = \tau_\varepsilon(D_p(m_p)), \quad \hat{D}_q(n_q) = \tau_\varepsilon(D_q(n_q)), \quad (2.18)$$

$$\tau_\varepsilon(x) = \begin{cases} x, & \text{if } x < \varepsilon, \\ \varepsilon, & \text{otherwise,} \end{cases} \quad (2.19)$$

where ε is a threshold proportional to the extrinsic calibration error of the system.

The function $V(m_p, m_u)$ is defined in a manner that preserves depth continuity between adjacent pixels. Choi and Lee [4] assume a pair of measured 3D points \mathbf{X}_p and \mathbf{X}_u to have been projected from close surface points if they are close to each other and have similar corrected amplitude values. The proximity is preserved by penalizing the pair of pixels if they have different numbers of wrappings:

$$V(m_p, m_u) = \begin{cases} \frac{\lambda}{r_{pu}} \exp\left(-\frac{\Delta\mathbf{X}_{pu}^2}{2\sigma_{\mathbf{X}}^2}\right) \exp\left(-\frac{\Delta A'_{pu}}{2\sigma_{A'}^2}\right) & \text{if } \begin{cases} m_p \neq m_u \text{ and} \\ \Delta\mathbf{X}_{pu} < 0.5d_{\max}(f_1) \end{cases} \\ 0 & \text{otherwise.} \end{cases}$$

where λ is a constant, $\Delta\mathbf{X}_{pu}^2 = \|\mathbf{X}_p - \mathbf{X}_u\|^2$, and $\Delta A'_{pu} = \|A'_p - A'_u\|^2$. The variances $\sigma_{\mathbf{X}}^2$ and $\sigma_{A'}^2$ are adaptively determined. The positive scalar r_{pu} is the image coordinate distance between p and u for attenuation of the effect of less adjacent pixels. The function $V(n_q, n_v)$ is defined by analogy with $V(m_p, m_u)$.

Choi and Lee [4] minimize the MRF energies via the α -expansion algorithm [2], obtaining a pair of unwrapped depth maps. To enforce further consistency between the unwrapped depth maps, they iteratively update the MRF energy corresponding to a depth map, using the unwrapped depth of the other map, and perform the minimization until the consistency no longer increases. Figure 2.10e, f shows examples of unwrapped depth maps, as obtained by the iterative optimizations. An alternative method for improving the depth accuracy using two ToF cameras is described in [3].

2.3.3 Discussion

Table 2.1 summarizes the phase-unwrapping methods [4–7, 14, 17, 20, 21] for ToF cameras. The last column of the table shows the extended maximum range, which can be theoretically achieved by the methods. The methods [6, 7, 14] based on the classical phase-unwrapping methods [9, 12] deliver the widest maximum range. In [4, 5], the maximum number of wrappings can be determined by the user. It follows that the maximum range of the methods can also become sufficiently wide, by setting N to a large value. In practice, however, the limited illuminating power of commercially available ToF cameras prevents distant objects from being precisely measured. This means that the phase values may be invalid, even if they can be unwrapped. In addition, the working environment may be physically confined. For the latter reason, Droschel et al. [6, 7] limit the maximum range to $2d_{\max}$.

2.4 Conclusions

Although the hardware system in [20] has not yet been established in commercially available ToF cameras, we believe that future ToF cameras will use such a frequency modulation technique for accurate and precise depth measurement. In addition, the phase-unwrapping methods in [4, 6] are ready to be applied to a pair of depth maps acquired by such future ToF cameras, for robust estimation of the unwrapped depth values. We believe that a suitable combination of hardware and software systems will extend the maximum ToF range, up to a limit imposed by the illuminating power of the device.

References

1. Bilmes, J.: A gentle tutorial of the EM algorithm and its application to parameter estimation for gaussian mixture and hidden Markov models. Technical Report TR-97-021, University of California. Berkeley (1998)
2. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell* **23**(11), 1222–1239 (2001)
3. Castañeda, V., Mateus, D., Navab, N.: Stereo time-of-flight. In: *Proceedings of the International Conference on Computer Vision (ICCV)*, pp. 1684–1691 (2011)
4. Choi, O., Lee, S.: Wide range stereo time-of-flight camera. In: *Proceedings of International Conference on Image Processing (ICIP)* (2012)
5. Choi, O., Lim, H., Kang, B., Kim, Y., Lee, K., Kim, J., Kim, C.: Range unfolding for time-of-flight depth cameras. In: *Proceedings of International Conference on Image Processing (ICIP)*, pp. 4189–4192 (2010)
6. Droschel, D., Holz, D., Behnke, S.: Multifrequency phase unwrapping for time-of-flight cameras. In: *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Taipei (2010)
7. Droschel, D., Holz, D., Behnke, S.: Probabilistic phase unwrapping for time-of-flight cameras. In: *Joint 41st International Symposium on Robotics and 6th German Conference on Robotics* (2010)
8. Fälie, D., Buzuloiu, V.: Wide range time of flight camera for outdoor surveillance. In: *Microwaves, Radar and Remote Sensing Symposium*, pp. 79–82 (2008)
9. Frey, B.J., Koetter, R., Petrovic, N.: Very loopy belief propagation for unwrapping phase images. In: *Advances in Neural Information Processing Systems* (2001)
10. Ghiglia, D.C., Romero, L.A.: Robust two-dimensional weighted and unweighted phase unwrapping that uses fast transforms and iterative methods. *J. Opt. Soc. Am. A* **11**(1), 107–117 (1994)
11. Göktürk, S.B., Yalcin, H., Bamji, C.: A time-of-flight depth sensor–system description, issues and solutions. In: *Proceedings of the Computer Vision and Parallel Recognition (CVPR) Workshops* (2004)
12. Goldstein, R.M., Zebker, H.A., Werner, C.L.: Satellite radar interferometry: two-dimensional phase unwrapping. *Radio Sci.* **23**, 713–720 (1988)
13. Jakowatz Jr, C., Wahl, D., Eichel, P., Ghiglia, D., Thompson, P.: *Spotlight-mode Synthetic Aperture Radar: A Signal Processing Approach*. Kluwer Academic Publishers, Boston (1996)
14. Jutzi, B.: Investigation on ambiguity unwrapping of range images. In: *International Archives of Photogrammetry and Remote Sensing Workshop on Laserscanning* (2009)
15. Liang, V., Lauterbur, P.: *Principles of Magnetic Resonance Imaging: A Signal Processing Perspective*. Wiley-IEEE Press, New York (1999)
16. Mesa Imaging AG. <http://www.mesa-imaging.ch>
17. McClure, S.H., Cree, M.J., Dorrington, A.A., Payne, A.D.: Resolving depth-measurement ambiguity with commercially available range imaging cameras. In: *Image Processing: Machine Vision Applications III* (2010)
18. Meyer, F.: Topographic distance and watershed lines. *Signal Process.* **38**(1), 113–125 (1994)
19. Opreşescu, Ş., Fälie, D., Ciuc, M., Buzuloiu, V.: Measurements with ToF cameras and their necessary corrections. In: *IEEE International Symposium on Signals, Circuits & Systems* (2007)
20. Payne, A.D., Jongenelen, A.P.P., Dorrington, A.A., Cree, M.J., Carnegie, D.A.: Multiple frequency range imaging to remove measurement ambiguity. In: *9th Conference on Optical 3-D, Measurement Techniques* (2009)
21. Poppinga, J., Birk, A.: A novel approach to efficient error correction for the swissranger time-of-flight 3D camera. In: *RoboCup 2008: Robot Soccer World Cup XII* (2008)
22. Rother, C., Kolmogorov, V., Blake, A.: “GrabCut”—interactive foreground extraction using iterated graph cuts. In: *International Conference and Exhibition on Computer Graphics and Interactive Techniques* (2004)