Simultaneous Time-of-Flight Sensing and Photometric Stereo with a single ToF Sensor

Changpeng Ti[1], Ruigang Yang[1], James Davis[2], Zhigeng Pan[3]
[1]University of Kentucky. [2]University of California, Santa Cruz. [3]Hangzhou Normal University.

Time-of-Flight (ToF) cameras based on phase-shift (such as PMD [3], SwissRanger [2]) take four snapshots to generate a metric depth map. In this paper, we develop novel techniques to allow Phase-shift-based ToF cameras to measure both metric distance and surface normal.

Our method uses four LED sources placed away from the imaging sensors, as shown in Figure 1. We illuminate the scene with these four lights, one at a time. Based on four phase images from such a distributed light setup, new algorithms are developed to recover both the surface normal and the distance for each pixel.

The classic ToF model requires that (1) the extra phase delay caused by the distance between the light source and the camera is negligible, and (2) pixels in the four phase images share the same intensity value. In our distributed light setup, neither condition remains true.

To tackle this problem, we explicitly model the phase delay caused by light positions as an unknown and use an iterative optimization scheme to solve it as well as the phase delay caused by scene depth.

One can use the classic formula to generate an initial phase delay estimate, and use the light source positions to perform further iterative refinements. In each iteration, a new phase delay ϕ' is estimated based on the current ϕ. Given the unit camera ray r, and the light positions L_i, where $i = \{0,1,2,3\}$ is the index of the light sources, one can write the phase delay α_i caused by the light source positions as

$$\begin{cases} d = \frac{c\cdot\phi}{4\pi\cdot f_{mod}} \\ \alpha_i = (||d\cdot r - L_i|| - d)\cdot\frac{2\pi\cdot f_{mod}}{c} \end{cases} \tag{1}$$

The relation between the phase delay ϕ' due to depth and the phase delay α due to light source positions can be written as

$$A\cos(\frac{i\pi}{2} + \alpha_i + \phi') = I_i = \frac{Im^i_{phase}}{Im^i_{intensity}} \tag{2}$$

where A is the amplitude of the ToF autocorrelation function (ACF), I is the value of the ACF, and Im_{phase} and $Im_{intensity}$ are the intensity values in the phase and intensity images respectively. This expands to (in matrix form)

$$\begin{bmatrix} \cos\alpha_i & -\sin\alpha_i \\ \cos(\frac{\pi}{2}+\alpha_i) & -\sin(\frac{\pi}{2}+\alpha_i) \\ \cos(\pi+\alpha_i) & -\sin(\pi+\alpha_i) \\ \cos(\frac{3\pi}{2}+\alpha_i) & -\sin(\frac{3\pi}{2}+\alpha_i) \end{bmatrix} \begin{bmatrix} A\cos\phi' \\ A\sin\phi' \end{bmatrix} = \begin{bmatrix} I_0 \\ I_1 \\ I_2 \\ I_3 \end{bmatrix} \tag{3}$$

$A\cos\phi$ and $A\sin\phi$ can be solved in the least-square sense in this overdetermined system. ϕ' is then estimated as

$$\phi' = \arctan(\frac{A\sin\phi'}{A\cos\phi'}) \tag{4}$$

The process from Eq. 1 to Eq. 4 is repeated until $||\phi' - \phi||$ is smaller than a threshold, or when the number of iterations reaches a limit. Upon convergence, we will obtain a refined depth map of the scene.

The Photometric Stereo method estimates the surface normal N by solving a system based on the Lambertian assumption. With point light sources, the light direction vectors need to be calculated for every pixel. Similar to the work by Clark [1], we further use a quadratic term to model light attenuation. Assuming the light source is infinitely small, the Lambertian shading equation can be rewritten as (in matrix form)

$$\begin{bmatrix} V^\top_{0,i} \\ V^\top_{1,i} \\ V^\top_{2,i} \\ V^\top_{3,i} \end{bmatrix} \cdot k_d N_i = \begin{bmatrix} I_0\cdot||L_0 - D_i||^2 \\ I_1\cdot||L_1 - D_i||^2 \\ I_2\cdot||L_2 - D_i||^2 \\ I_3\cdot||L_3 - D_i||^2 \end{bmatrix} \tag{5}$$

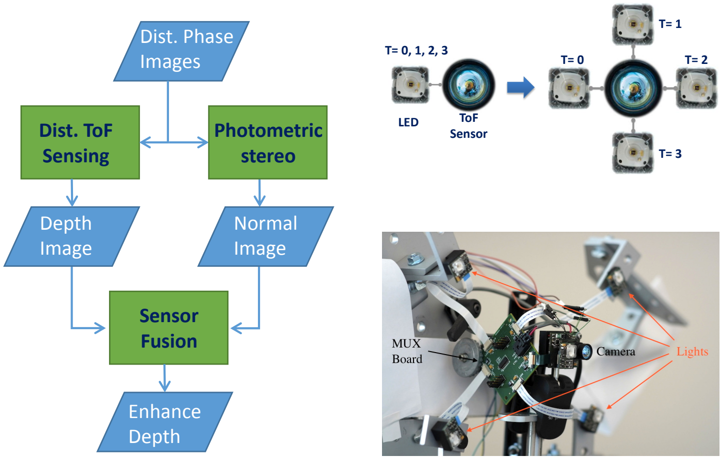

Figure 1: (Left) overview of our depth enhancenemt framework. (Right) illustrations of our novel capturing setup that consists of a ToF sensor and four point light sources.

where $V_{k,i} = L_k - D_i$ represents light direction from the k-th light to the i-th pixel, and D_i represents the 3D coordinates estimated by Eq. 4.

The position information from range map and the surface normal from Photometric Stereo is combined using a framework similar to [4]. In their approach, they iterative refine the position and the normal. During the position refinement step, the following objective function is minimized

$$\begin{aligned} E \;=\; & \sum_{ij}\mu_{ij}(D_{ij} - D^0_{ij})^2 + \lambda_p \sum_{ij}\mu_{ij}\Big((N^0_{ij}\cdot\frac{\partial D}{\partial u})^2 + \\ & (N^0_{ij}\cdot\frac{\partial D}{\partial v})^2\Big) + \lambda_s \nabla D \end{aligned}$$

where D^0_{ij} and N^0_{ij} are the depth and normal, which are obtained via distributed ToF sensing (Eqs. 1∼4) and Photometric Stereo (Eq. 5). $\{\mu_{ij}\}$ are per-pixel weights and $\frac{\partial D}{\partial u}$ and $\frac{\partial D}{\partial v}$ are derivatives of the depth map along the image grid.

With small and low-cost changes to the hardware, our sensor fusion pipeline leads to much improved depth map in terms of metric accuracy and surface details. We demonstrate the capability of our novel hardware/software system with a number of quantitative and qualitative experiments. Looking into the future we hope to explore a multi-light multi-ToF-sensor setup to further improve the quality of the depth map and allow the system to scale up.

[1] James J Clark. Active photometric stereo. In *Computer Vision and Pattern Recognition, 1992. Proceedings CVPR'92., 1992 IEEE Computer Society Conference on*, pages 29–34. IEEE, 1992.

[2] T. Oggier, B. Büttgen, F. Lustenberger, G. Becker, B. Rüegg, and A. Hodac. Swissranger sr3000 and first experiences based on miniaturized 3D-ToF cameras. In *Proc. of the First Range Imaging Research Day at ETH Zurich*, 2005.

[3] Z. Xu, R. Schwarte, H. Heinol, B. Buxbaum, and T. Ringbeck. Smart pixel – photonic mixer device (PMD). In *Proc. Int. Conf. on Mechatron. & Machine Vision*, pages 259–264, 1998.

[4] Qing Zhang, Mao Ye, Ruigang Yang, Yasuyuki Matsushita, Bennett Wilburn, and Huimin Yu. Edge-preserving photometric stereo via depth fusion. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2472–2479. IEEE, 2012.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.