Fusion of Time-of-Flight Depth and Stereo for High Accuracy Depth Maps

Jiejie Zhu[†] Liang Wang[†] Ruigang Yang[†] James Davis[‡]

[†]Center for Visualization and Virtual Environments, University of Kentucky, USA
[‡]Computer Science Department, University of California, Santa Cruz, USA

Abstract

Time-of-flight range sensors have error characteristics which are complementary to passive stereo. They provide real time depth estimates in conditions where passive stereo does not work well, such as on white walls. In contrast, these sensors are noisy and often perform poorly on the textured scenes for which stereo excels. We introduce a method for combining the results from both methods that performs better than either alone. A depth probability distribution function from each method is calculated and then merged. In addition, stereo methods have long used global methods such as belief propagation and graph cuts to improve results, and we apply these methods to this sensor. Since time-of-flight devices have primarily been used as individual sensors, they are typically poorly calibrated. We introduce a method that substantially improves upon the manufacturer's calibration. We show that these techniques lead to improved accuracy and robustness.

1. Introduction

Depth sensing is one of the fundamental challenges of computer vision. Applications include robotic navigation, object reconstruction, and human computer interaction. A range sensor that is robust, accurate, and real time would be the enabling component in these applications. Unfortunately no existing range sensing method is perfect on its own. Laser scanners are too slow for real time use. Passive stereo fails on textureless scenes. Time-of-flight sensors are low resolution, noisy, and poorly calibrated. Photometric stereo is prone to low frequency distortions.

Time-of-flight sensors provide real time independent range estimates at each pixel, and are only recently becoming available from companies such as Canesta [2], SwissRanger [3], and 3DV [1] at commodity prices. Due to their recent introduction, most applications use the sensors individually and rely on the manufacturer's calibration. Despite their promise, relatively little literature explores the ways in which the quality of these sensors might be improved.

This paper seeks to improve the range estimates provided by time-of-flight sensors, by combining them with both the data and relatively more sophisticated algorithms common to passive stereo vision.

Time-of-flight sensors are characterized by independent pixel range estimates, each of which has relatively high noise that can be modeled as a Poisson distribution around the true depth. Texture causes difficulties since these sensors frequently have biases as a function of object albedo. Passive stereo is characterized by outliers and correlations between neighboring pixels. Textureless regions and repeated patterns will cause multiple local minima to appear in the cost volume, resulting in a relatively complex error model.

We combine the probability distribution functions on depth from each of these sensor modalities using a Markov random field (MRF) model to produce a combined sensor with superior characteristics.

Modern passive stereo methods are quite sophisticated and have been carefully categorized according to the effects of changing the local matching function, aggregation function, and global regularization [22]. Time-of-flight sensors provide only local data, analogous to performing stereo using a simple sum-of-squared-distance (SSD) correspondence match only. We apply belief propagation (BP) as a global regularization term and show that it improves time-of-flight accuracy, just as it improves stereo.

Any attempt to combine multiple sensors requires that they be calibrated in a common coordinate frame. Unfortunately time-of-flight sensors are usually designed only to provide relative depth, as opposed to measurements in a calibrated Euclidean frame.

We introduce an empirical calibration method which builds a look-up table (LUT) with 4 dimensions, mapping observed intensity and 3D positions reported by the sensor to ground truth distance. Table entries are obtained by moving a planar card to known depths in a 400 mm calibration region. We quantize intensity into simply black and white, linearly interpolating for values in between. Results show the depth error within the calibrated distance range is reduced from 1.8% to 0.6% of the working volume.

978-1-4244-2243-2/08/$25.00 ©2008 IEEE

The contributions of this paper include: a method for using data from both time-of-flight and stereo methods to produce enhanced depth estimates, application of global regularization methods to time-of-flight data, and a method for improving calibration of time-of-flight sensors.

2. Related Works

There are many ways to obtain scene depth information. In general, they can be categorized into two major classes: passive range sensors and active range sensors. Among the plethora of passive range sensing methods, stereovision [23] is probably the most well known, least expensive, and most widely used. It is beyond the scope of this paper to provide even a brief review of existing stereo methods, interested readers are refereed to an excellent review by Scharstein and Szeliski [22]. Despite significant progress made during the last few years, the fundamental problems in stereo such as occlusion and textureless regions, remain to be unsolved. Thus, stereo results from real scenes are often quite fragile.

Time-of-Flight (TOF) range sensors use an active technique to obtain near real-time scene depth. The basic principle of depth calculation in TOF is based on timing the round-trip of a pulse of light. TOF sensors are able to produce a full depth frame simultaneously, therefore allowing their application to dynamic scenes. Lange [21] describes a TOF camera using modulated visible and near-infrared radiation. An excellent classification of TOF for acquiring depth can be found in [17]. A key issue for TOF depth sensors is low spatial resolution, making these sensors less appealing for most vision algorithms. Therefore several super resolution methods [10] [26] are reported to enhance the resolution of depth images. These are mainly based on the fact that discontinuities in range and coloring tend to coalign [12]. As a separate issue, the capture frequency can be increased by a CMOS-based technique [6]. Several types of commercial TOF sensors are now available [3, 2, 1].

Although the basic principle of TOF sensors is simple, the accuracy of depth measurement is subject to many factors. The main noise source is photon shot noise which is theoretically Poisson distributed [5]. Others, like flicker noise and thermal noise, can be modeled as a constant. Anisotropic LED lights and lens distortions are regarded as internal errors which can be reduced by calibration [14]. External uncertainties like environmental temperature, background light, multiple reflection, light scattering, glossy reflection and color difference all have impacts on accuracy. Ambient light is particularly important and special purpose calibration [11] as well as in-pixel background light suppression [17] have been investigated. However, there is currently no general and simple methods to handle general noise and most applications use depth from TOF sensors directly, which may not be sufficient in high accuracy applications.

There are already several approaches to merge a TOF sensor with a stereo rig. A comparisons between depth from TOF and depth from stereo rig has also been performed [8]. Refined 3D estimation by regional selection of depth from a stereo rig and depth from TOF is reported by [16, 20, 7].A similar setup to ours is [4] which also improves the 3D estimation by finding the correspondence between images from TOF sensor and stereo rig. However, none of these methods reported metric improvement over TOF or stereo sensors, nor provided disparity images comparable in quality to the current state of the art in stereo. In this paper we present a general calibration method that does not assume specifics of our TOF sensor and provide a detailed evaluation on a number of real scenes.

3. Multi-Sensor Setup

We combine a TOF sensor with a pair of stereo cameras (as shown in Figure 1). The TOF sensor we have is a Swiss Ranger SR3000 [3], which can continuously produce a depth map of 176×144 resolution with an operational range up to seven meters. In our current setup, the two CCD stereo cameras have a baseline about 100mm and are verged towards each other at five degree from the parallel setup. Our set up is designed to provide optimal coverage at about one meter range.

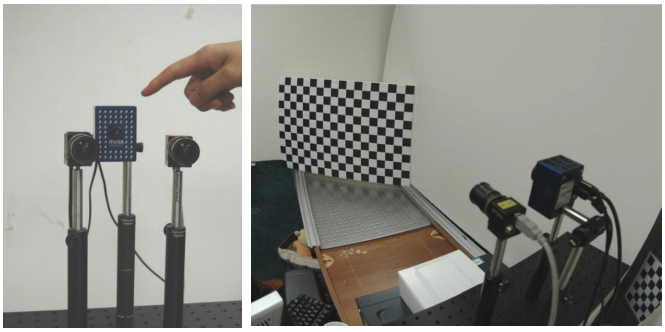


Figure 1. Multi-sensor setup(Left). Calibration setup, notice the rails on the table to move the pattern(Right).

4. Multi-Sensor Calibration

Our calibration method utilizes the stereo cameras to triangulate 3D points (X_{st}) as ground truth to guide all steps. Since stereo camera calibration is a well studied problem we do not need to construct elaborate 3D calibration objects for any sensor. We limit the calibration range from 1m to 1.40m using 16 distance intervals which we called steps (each step has a distance size around 25mm). To minimize saturation effects, optimal camera integration time procedure [18] is used.

4.1. Geometric Calibration

During our calibration procedure, a planar object covering the largest field of view is vertically placed to face the stereo and sensors. This plane, painted with checkerboard patterns, is movable on a metric board with two guide rails (shown in Figure 1 right). We move the pattern from near to far in each step, note that precise movement is not necessary.

In each step, stereo captures the left/right view while the TOF sensor returns an intensity image and a depth image with (x, y, z) information. The correspondence between the stereo image and sensor depth image is defined by a homographical transformation. There are altogether three coordinate systems (CS) involved in our geometrical model: the stereo CS, the calibration (world) CS and the sensor CS. 3D points in these CS are denoted as $X_{\text{st}}, X_{\text{dc}}$, and X_{ds}, respectively. The theoretical explanation of their relationship is illustrated in Figure 2:

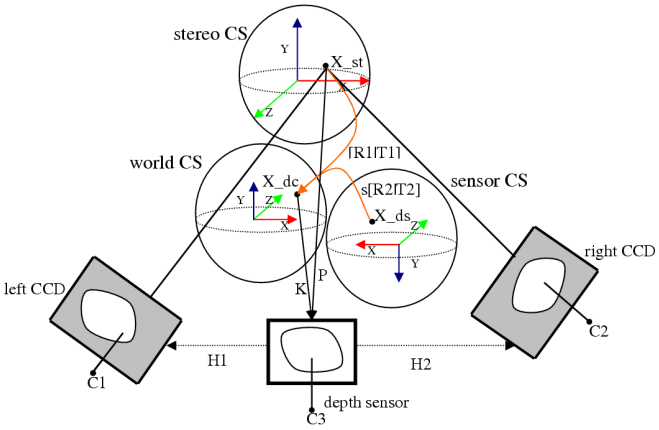


Figure 2. Geometric relationship of depth sensor and stereo rig.Since the scene is a plane (for calibration), the geometrical relationship between the depth sensor and the left/right camera is homograph $(H1, H2)$. From standard camera calibration, we can estimate $[R1|T1]$ to transform between X_{st} and X_{dc}. And assuming rigid transformation between X_{dc} and X_{ds}, the geometrical relationship among three CSs are fixed. K, P is the intrinsic matrix and projective matrix of sensor.

Referring to Figure 2, The transformation between the three cameras (treating the TOF's intensity image as from a camera) can be readily computed using standard camera calibration techniques (such as [28]). Besides, we know the correspondence between X_{ds} and its 2D location p_{ds} on the sensor's intensity image, however we do not know the "back-projection" to map the 2D pixel to X_{ds}, causing problem to transform X_{ds} to X_{dc} for correction and fusion.

Given a pixel p_{ds} from the sensor, we find its ground truth in calibration using the following steps:

1. Calculate matched pair in the stereo rig by H_1 and H_2.
2. Triangulate the matched pair and get its 3D point X_{st} in stereo CS.
3. Transform from stereo CS to calibrated CS using $[R_1|T_1]$, and we get the ground truth X_{dc}.

Now we have two coordinates for p_{ds}, X_{dc} and X_{ds}. This 3D-3D correspondences can be used to estimate the transformation $s[R_2T_2]$ by a closed form method [19]. Thus, the alignment matrix between X_{st} and X_{ds} is:

$$\bar{X}_{\text{ds}} = s[R_2T_2]^{-1}[R_1T_1]X_{st} = \mathbf{M}X_{st} \tag{1}$$

Based on above analysis, we calculate the distance correction value between X_{ds} and X_{st} in pixel simply by:

$$X_d = \bar{X}_{\text{ds}} - X_{\text{ds}} = \mathbf{M}X_{\text{st}} - X_{\text{ds}} \tag{2}$$

In each step, a per pixel array of distance correction value is calculated and added to a distance look up table (LUT). After calibration, we have a 3D table in which each cell stores the correction value X_d. During run-time, we can use this LUT to refine sensor depth. Given X_{ds}, we first locate two nearest cells from the LUT through X_{ds}'s sensor pixel coordinates and reported sensor range. Through linear interpolation, the distance correction for X_{ds} is calculated and applied to X_{ds}.

4.2. Photometric Calibration

Based on the fact that the mean accuracy correlates with mean intensity of the scene [18], white and black planar object are used to calibrate the intensity influence on distance accuracy. Figure 3 shows depth value both in white and black intensity in 16 steps. Only the result of the 72th scan line is plotted for clarity. We see that an almost constant shift exists between the white and black in each step. Based on this fact, we build a per-pixel intensity-range bias table,

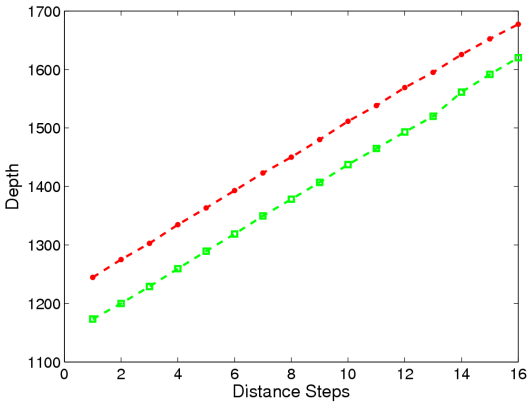


Figure 3. Distance of the 72th scan line from white and black. In each distance step (25mm), the distance difference is quite same (within 5mm), which shows that intensity influence on depth accuracy is independent of distance.

which basically record the constant range bias between a black and a white reference object. For a gray-scale object, the range bias is proportionally adjusted based on its intensity difference to the black and white reference value. For each TOF sensor image, we first perform this photometric-based correction, then apply the geometric correction in Section 4.1.

4.3. Calibration Verification

To verify our calibration results, plane and box are tested. Depth error and pixel re-projection error (PRE) on three views are analyzed. PRE is defined as the mean difference between 3D points re-projected on a view and their positions originally selected manually.

The Plane Experiment In this test, a planar board with checker patterns is placed inside the calibration range. Data are captured in 5 postures: orthogonal, forward, backward, left and right (see figure 4).

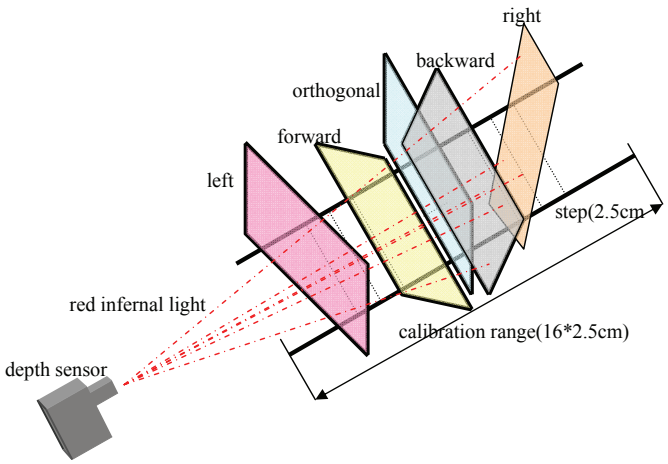
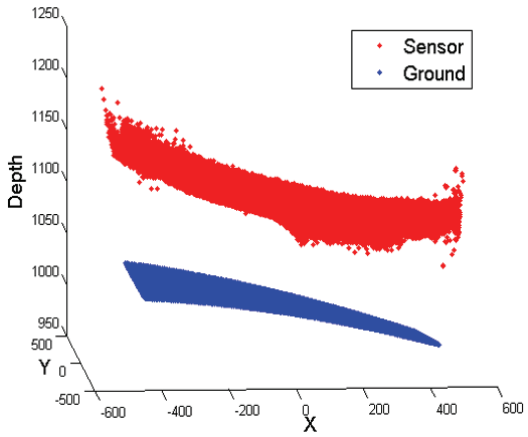


Figure 4. LUT verification using planar plane. 5 postures are tested, and each is placed at different position but within calibration range.

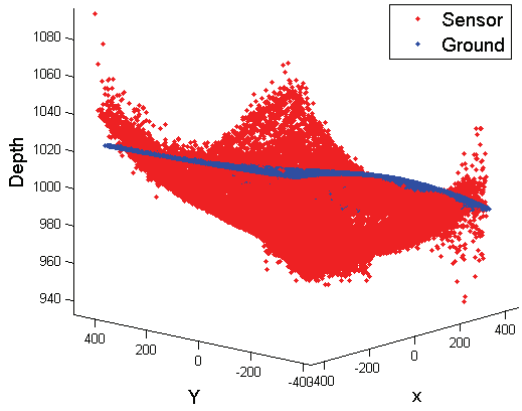
Figure 5 shows the effect of the rigid transformation. One plane posture is used. Figure 6 plot the 3D locations of depth points from different methods. One can see that the ones after LUT refinement are much closer to the ground truth.

PRE and mean metric depth error (e.g.,X_d) are presented in Table 1. This table shows points after LUT refinement have much smaller error, which is approximately one third of their original value.

The Box Experiment In this test, we place a box inside the calibration range and let two orthogonal patches be the target. We want to show that the orthogonal relationship is better preserved by our LUT method than the TOF sensor.



(a)



(b)

Figure 5. Visual comparison of rigid transforation. In (a), raw data from sensor (in red color) is obviously not in the same coordinate system with ground truth (in blue color). After rigid transformation , it is aligned in world CS in (b).

As same in first test, we read out raw points from sensor and calculated revised points by LUT. To compare the metric accuracy, the angle between two patches is estimated using 3D plane fitting method. Figure 7 illustrate the results.

Based on above tests, we verify that our calibration LUT improves the depth accuracy. To further improve the depth accuracy, we present our fusion technique in the following section.

5. Sensor Fusion

The current state of the art in stereo matching is achieved by global optimization algorithms (e.g., [15, 13, 25]). These methods formulate stereo matching as a maximum a posterior Markov Random Fields (MAP-MRF) problem. In detail, we denote $X = x_i$ as hidden variables, corresponding to the disparities of each pixel and $Y = y_i$ as observed vari-

Table 1. Result of depth error (in the second column only) and re-projection error (in pixel). Numbers in **bold** are after LUT refinement. The re-projection error is much smaller on the sensor view because of its low resolution.

Posture	Depth Error (mm)	Pixel Error Sensor View	Pixel Error Left View	Pixel Error Right View
orthogonal	17.0 (5.3)	(0.34 0.36) (0.02 0.04)	(1.82 2.35) **(1.85 1.63)**	(3.50 2.21) **(1.91 1.72)**
left	17.0 (6.1)	(0.35 0.28) (0.03 0.06)	(3.33 2.68) **(1.87 1.64)**	(1.68 2.62) **(2.44 1.65)**
right	20.3 (5.0)	(0.32 0.35) (0.02 0.05)	(1.96 2.44) **(1.77 1.67)**	(4.21 2.23) **(2.91 1.65)**
forward	19.2 (5.1)	(0.39 0.31) (0.02 0.08)	(6.24 4.72) **(3.75 3.33)**	(5.81 4.60) **(4.41 3.14)**
backward	16.3 (4.6)	(0.36 0.33) (0.02 0.06)	(2.44 2.48) **(1.86 1.69)**	(3.12 2.31) **(2.10 1.71)**

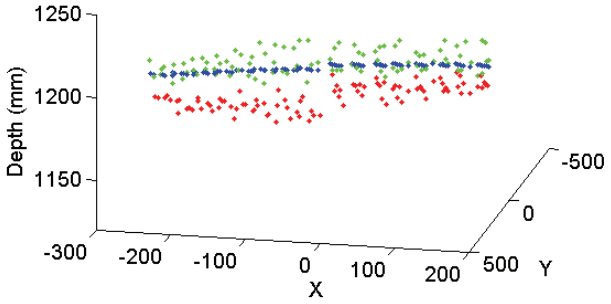


Figure 6. Visual comparison of points before and after calibration. Ground truth is plotted in blue, points after LUT correction is in green, and point after rigid transformation is in red. The view direction is aligned with the ground truth plane.

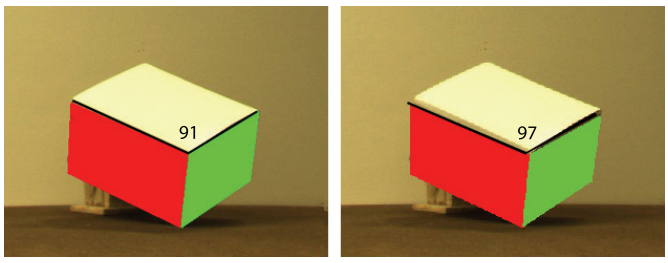


Figure 7. Angle comparison of two orthogonal patches. Using 3D plane fitting, the angle from LUT refined method is 91° while that from the sensor is 97°.

ables, corresponding to the intensity based matching cost at specific disparity. Solving the stereo matching problem is equivalent to maximize the following posterior.

$$P(X|Y) \propto \prod_i f_d(x_i, y_i) \prod_i \prod_{j \in N(i)} f_s(x_i, x_j) \quad (3)$$

where $N(i)$ represents the neighbors of node i, function f_d is the local evidence for node i based on the initial pixel-wise matching cost (*data term*) and f_s is a symmetric function measures the smoothness assumption about the scene. One nice feature of this MAP-MRF formulation is that it provides a natural way to integrate the information from multiple sensors. With our TOF sensor, we define a set of observed variables $Z = z_i$, which corresponds to the depth value returned by the sensor. The new posterior is

$$P(X|Y,Z) \propto \prod_i f_d(x_i, y_i) f_r(x_i, z_i) \prod_i \prod_{j \in N(i)} f_s(x_i, x_j). \quad (4)$$

where $f_r(x_i, z_i)$ is the additional local evidence based on the measurement from the TOF sensor.

We choose belief propagation (BP) to maximizing $P(X|Y,Z))$, or in fact $\log(P(X|Y,Z))$. We also introduce two weighting factors to allow more flexibility for the data term (after log). That is

$$d = w_s \cdot \log f_d(x_i, y_i) + w_r \cdot \log f_r(x_i, z_i) \quad (5)$$

where w_d and w_r are the weighting factors for stereo and TOF data term. By setting w_r to zero, we have a standard BP-based stereo algorithm; and by setting w_d to zero, we have an algorithm to globally optimize the raw sensor depth data. While simple in retrospect, we do not believe such a global-optimization based approach has been applied to any of these active sensors, such as structured light or TOF sensors.

5.1. Data Term from Stereo Matching

The data term derived from stereo matching encodes the color consistency of pixel correspondences. In our implementation the pixelwise matching cost are obtained using a similar manner as was recently proposed by [24]. In detail, the per-pixel difference is first computed using Birchfield and Tomasi's pixel dissimilarity [9] and an additional adaptive weight aggregation step [27] is applied to overcome matching ambiguities caused by occlusion boundaries or sensor noise *etc*. This two-step approach has shown to be remarkably effective for getting reliable matching cost in [24].

5.2. Data Term from the TOF Sensor

The data term from depth sensor encodes the depth consistency between the stereo and depth sensor. The cost volume is built based on the current depth map using quadratic error model. In order to allow large depth variations, as current depth values from depth sensor are not guaranteed to be correct, the cost function should become constant as the differences become large. Given a range of disparity candidates, the cost volume is computed by following steps:

firstly we triangulate all potential correspondences and get 3D points candidates X_{st}^{c}; secondly we transform X_{st}^{c} to the calibrated CS X_{dc}^{c1} and project them onto the intensity image with result of 2D pixel candidates uv_{dc}^{c}; thirdly, we select the 3D points in the sensor CS X_{ds}^{c} based on uv_{dc}^{c} and transform them to the calibration space by LUTs X_{dc}^{c2}; finally, the depth cost is calculated based on the Z difference between X_{dc}^{c1} and X_{dc}^{c2}.

$$f_r = min(\eta * L, (Z_{\mathrm{dc}}^{c1} - Z_{\mathrm{dc}}^{c2})^2) \tag{6}$$

L is the search range, η is a constant.

The two weighting factors are empirically calculated by comparing with ground truth (See Section 6). We notice that setting $w_s = 0.1$ and $w_r = 0.4$ achieves good results in most cases.

6. Experimental Results

In order to verify the accuracy of our methods, we set up a single structured light scanner to acquire scene depth. Basically we use a projector to project a single line sweeping over the scene. The orientation of the line is roughly orthogonal to the epipolar line of the cameras, therefore the correspondence problem can be uniquely determined in the stereoscopic image pair. Based on the cameras' calibration data and sensor resolution, our structured light scanner is able to achieve a depth accuracy of 5mm over a one-meter range.

In order to provide a fair comparison, we differentiate between local and global methods. This distinction was originally from stereo-vision. As defined in the taxonomy by Scharstein and Szeliski [22], a local method associates a pixel's disparity (depth) value to the one with the minimum matching cost, e.g., a local "winner-takes-all" (WTA) approach. In contrast, a global method typically makes disparity decision using an energy-minimization framework. We can then define the following different methods to compute the depth map using the range sensor alone: (A) local method, (B) local method with LUT correction, (C) global method, and (D) global method with LUT correction. (A) is essentially the raw depth reading reported by the depth sensor. To generate the depth map from (C), we just need to change the data term in the BP optimization framework by setting w_s in equation 5 to zero. (D) is similarly defined. The methods for using the stereo cameras alone include a local method (E) and a global method (F) that sets w_r in equation 5 to zero. Finally we can define a local fusion method (G), which applies WTA to the data term in equation 5, and a global fusion method (H) which applies the entire BP-based optimization.

We test our algorithms on a number of scenes. For each scene, we first use the scanner to obtain the ground truth, then apply methods (A)–(H) to compute the depth maps. We first show a simple scene (wall) with two planes: one is uniformly colored while the other has a checker board pattern. As expected, the TOF sensor and the stereo method exhibit complimentary characteristics for these two planes: the TOF sensor performs well in area with constant/smooth intensities while the stereo method performs well in area with a lot of textures. It is also interesting to see that the sensor's depth maps (in (C) and (D)) improved after applying a global optimization step.

Then we tested with two complex scenes (teddy and head). The results from global method are shown in Figure 9. Overall, fusion reduces the depth map's noise and provides better results on depth discontinuities.

Finally we tested with a scene (cup) containing a glossy object (Figure 10). We can see that the reading from TOF sensor is grossly wrong on the right side of the cup. It is due to inter-reflection; notice that the right side of the cup is brighter than the left side while the scene is uniformly illuminated from the top. BP-based stereo matching, which assumes a lambertian object, is in fact quite robust on objects with small specular highlights or inter-reflections. From this experiment we can see that our TOF sensor is very susceptible to inter-reflection. This is a place where sensor fusion can help substantially.

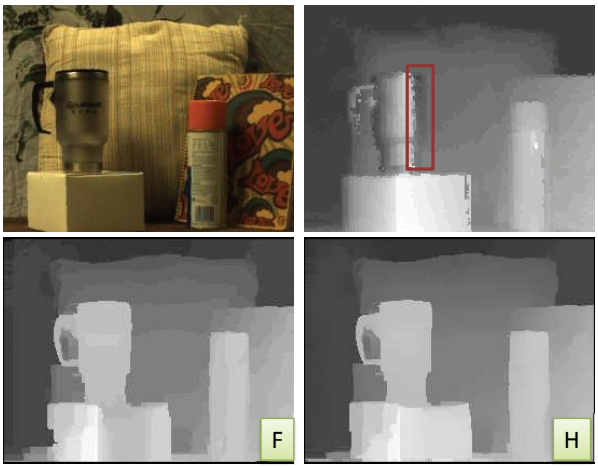


Figure 10. An test case with inter-reflection. The top right image is the raw sensor depth map. The depth in the marked area is wrong. (F) is from global BP-stereo alone and (H) is the globally fused result.

The numerical comparison against the ground truth is presented in Table 2. The last scene (cup) has strong specular reflection that distorted the result from our simple structured light scanner therefore it is not included for numerical evaluation. The fused result is always the best, reducing the distance error by half on average as compared to the raw sensor data (after rigid transformation). Comparing Table 2 with 1, we can see that the accuracy from real-world scenes is not as good as that from the controlled setup. We believe

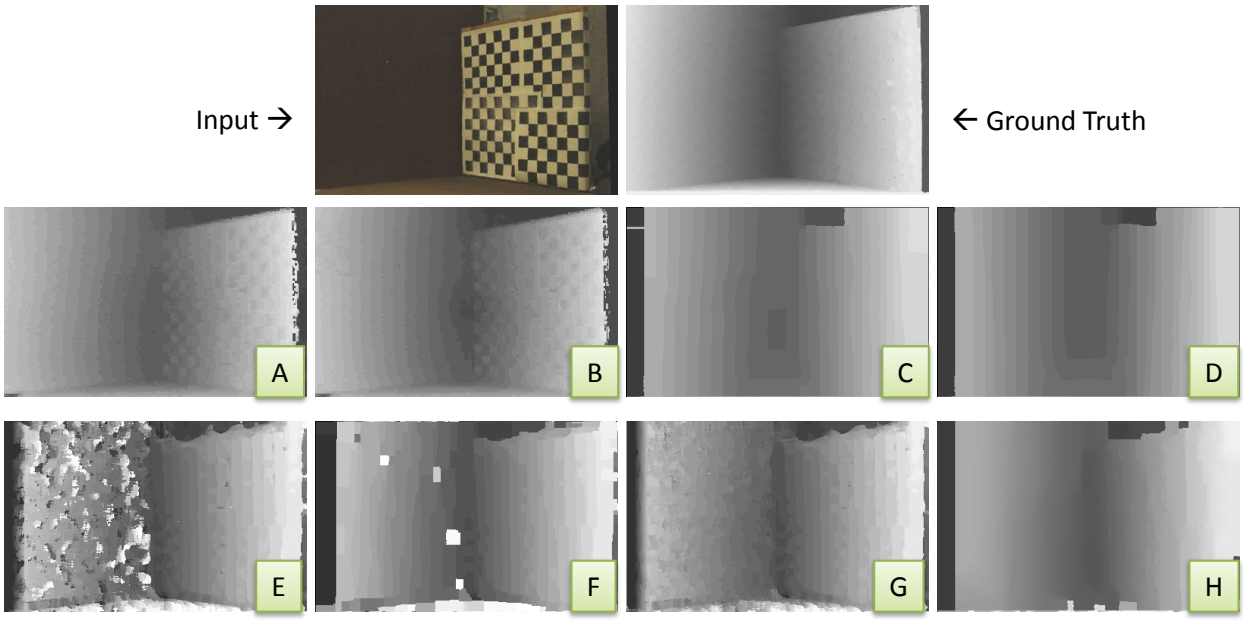


Figure 8. Depth map from a simple scene with two walls. Label A to H indicates which reconstruction method is used. (A) and (B) are calibrated depth maps from the sensor, (C) and (D) are after global optimization. (E) and (F) are the local and global stereo method, and (G) and (H) are from local and global fusion.

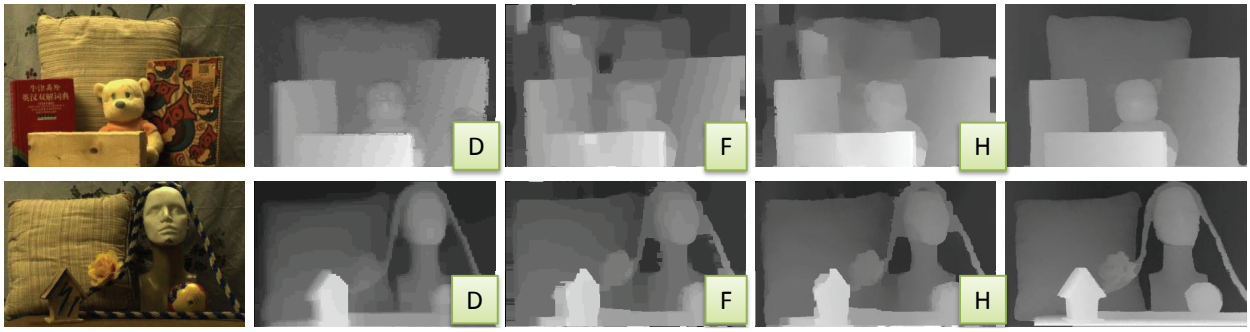


Figure 9. Depth maps from two complex scenes. The last column shows the ground truth. Global senor method (D), global stereo method (F), and global fusion method (H) are used.

it is due to the complex lighting and surface reflectance and texture variations. This is a place for further study.

7. Conclusion

In this paper we present a simple and effective calibration method to improve the performance of time-of-flight (TOF) sensors. Our method is general and needs no additional equipment other than a pair of cameras. We in fact use the stereo camera to build a virtual 3D calibration object so that the sensor can be calibrated in any desired range. Evaluation shows that our calibrated depth map can achieve an absolute accuracy of about 5mm over a range of one meter.

This is a about three times improvement compared to the raw depth map (after rigid transformation). We further present a probabilistic framework to fuse the depth maps from stereo and the TOF sensor. We show that for complex scenes with specular highlight, large texture variations, or even inter-reflections, our fused results reduce the over-all error by 50%.

In the future, we would like to explore the fusion of depth maps over dynamic scenes. We expect to see more dramatic improvement since TOF sensors are known to be sensitive to motion.

Acknowledgement This work is supported in part by US National Science Foundation grant IIS-0448185, a grant from US Department of Homeland Security, and a grant from HP lab.

Table 2. Numerical comparison with ground truth. Mean depth error (in millimeter) for scenes wall, teddy, and head; The number in *italic* is the mean disparity error.

	scene: wall		scene: teddy		scene: head	
	Local	Global	Local	Global	Local	Global
raw sensor	31.7 (*2.0*)	29.3 (*1.9*)	53.5 (*2.7*)	53.97 (*2.7*)	35.2 (*1.75*)	45.1 (*2.4*)
LUT corrected	49.7 (*2.85*)	23.0 (*1.2*)	27.0 (*1.3*)	28.8 (*1.3*)	40.6 (*2.0*)	42.1 (*2.2*)
stereo	41.2 (*2.4*)	33.3 (*1.9*)	36.8 (*1.9*)	22.0 (*1.0*)	35.5 (*1.83*)	28.5 (*1.4*)
fusion	28.6 (*1.7*)	19.0 (*1.0*)	21.6 (*1.0*)	17.1 (*0.8*)	28.7 (*1.4*)	24.0 (*1.2*)

References

[1] 3dv systems, z-cam. http://www.3dvsystems.com, 2004.

[2] Canesta inc, canestavision electronic perception development kit. http://www.canesta.com/, 2006.

[3] Swissranger inc, sr-2. http://www.csem.ch/fs/imaging.htm, 2006.

[4] S. A.Gumundsson, H. Aans, and R.Larsen. Fusion of stereo vision and time-of-flight imaging for improved 3d estimation. In *International workshop on Dynamic 3D Imaging*, 2007.

[5] S. A.Gumundsson, H.Aans, and R.Larsen. Environmental effects on measurement uncertainties of time-of-flight cameras. In *International Symposium on Signals, Circuits and Systems*, pages 1–4, 2007.

[6] A.H.Izhal, T.Ushinaga, T.Sawada, M. Homma, Y. Maeda, and S.Kawahito. A cmos time-of-flight range image sensor with gates-on-field-oxide structure. *IEEE Sensors Journal*, 7(12):1578–1586, 2007.

[7] T. D. A.Prasad, K.Hartmann, W.Weihs, S. E. Ghobadi, and A.Sluiter. First steps in enhancing 3d vision technique using 2d/3d sensors. In *Computer Vision Winter Workshop*, pages 82–86, 2006.

[8] C. Beder, B. Bartczak, and R. Koch. A comparison of pmd-cameras and stereo-vision for the task of surface reconstruction using patchlets. In *CVPR*, 2007.

[9] S. Birchfield and C. Tomasi. A pixel dissimilarity measure that is insensitive to image sampling. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20(4):401–406, 1998.

[10] S. B. Gktrk, H. Yalcin, and C. Bamji. A time-of-flight depth sensorsystem description, issues, and solutions. In *CVPR*, page 35, 2004.

[11] H. Gonzales-Banos and J. Davis. Computing depth under ambient illumination using multi-shuttered light. In *CVPR*, pages 234–241, 2004.

[12] J.Diebel and S.Thrun. An application of markov random fields to range sensing. In *NIPS*, pages 291–298, 2005.

[13] J.Sun, Y.Li, S. B. Kang, and H. Y. Shum. Symmetric stereo matcing for occlusion handling. In *CVPR*, pages 399–406, 2005.

[14] T. Kahlmann, F. Remondino, and H. Ingensand. Calibration of the fast range imaging camera swissranger for use in the surveillance of the environment. In *Proceedings of SPIE:Electro-Optical Remote Sensing II*, volume 6396, 2006.

[15] V. Kolmogorov and R. Zabih. Computing visual correspondence with occlusions using graph cuts. In *ICCV*, pages 508–515, 2001.

[16] K. D. Kuhnert and M. Stommel. Fusion of stereo-camera and pmd-camera data for real-time suited precise 3d environment reconstruction. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4780–4785, 2006.

[17] M. Lehmann, R. Kaufmann, F. Lustenberger, B. Bttgen, and T. Oggier. Ccd/cmos lock-in pixel for range imaging: Challenges, limitations and state-of-the-art. In *CSEM, Swiss Center for Electronics and Microtechnology.*, 2004.

[18] S. May, B.Werner, H. Surmann, and K. Pervlz. 3d time-of-flight cameras for mobile robotics. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1578–1586, 2006.

[19] B. K. P.Horn. Closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America*, 4(4):629–642, 1987.

[20] R. Reulke. Combination of distance data with high resolution images. In *ISPRS Commission V Symposium Image Engineering and Vision Metrology*, 2006.

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

[22] D. Scharstein and R. Szeliski. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *International Journal of Computer Vision*, 47(1):7–42, May 2002.

[23] A. Verri and V.Torre. Absolute depth estimate in stereopsis. *The Optical Society of America A*, 3:297–299, 1986.

[24] Q. Yang, L. Wang, R. G.Yang, H. Stewenius, and D. Nister. Stereo matching with color-weighted correlation, hierarchical belief propagation and occlusion handling. In *CVPR*, pages 2347–2354, 2006.

[25] Q. X. Yang, L.Wang, R. G.Yang, S. G.Wang, M.Liao, and D.Nister. Real-time global stereo matching using hierarchical belief propagation. In *BMVC*, 2006.

[26] Q. X. Yang, R. G. Yang, J.Davis, and D.Nister. Spatial-depth super resolution for range images. In *CVPR*, 2007.

[27] K. J. Yoon and I. S. Kweon. Locally adaptive support-weight approach for visual correspondence search. In *CVPR*, pages 924–931, 2005.

[28] Z. Zhang. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22:1330–1334, 2000.