Combination of Time-of-Flight Depth and Stereo using Semiglobal Optimization

Jan Fischer and Georg Arbeiter and Alexander Verl

***Abstract*— A growing number of modern computer vision applications like object recognition, collision avoidance and scene understanding demand accurate and dense 3D representations of their environment. To improve existing procedures for 3D data acquisition this paper proposes a novel method for sensor combination on a stereo and a Time-of-Flight camera system. By calibrating the two sensor systems to each other, valid measurements from the 2.5D Time-of-Flight sensor are converted to disparity guesses within the stereo system. The disparity guesses from the Time-of-Flight data constrain the correspondence search results from the stereo matching algorithm. It is empirically shown, that the proposed method effectively enhance the results from stereo vision, especially in structureless areas where stereo correspondence search fails. The method is evaluated on the camera system of the service robot Care-O-bot® 3.**

I. INTRODUCTION

The combination of sensor data from different sources aims at creating information that exceeds the quality of each individual source. In terms of quality one usually relates to accuracy, completeness or confidence. The data sources considered in this project are two color cameras used for stereoscopic vision and one Time-of-Flight sensor that directly delivers 2.5D range data. The paper aims at combining both modalities to create accurate 3D point clouds with associated color information even in unstructured image areas. In the following, the characteristics of the two sensor modalities are described, beginning with the Time-of-Flight sensor.

Time-of-Flight cameras emit modulated near infra-red light to illuminate a given scene. The reflection of the modulated light is collected in a CMOS matrix. By comparing the returning signal to the camera's source modulation, the phase shift is measured which is a linear function of distance to the reflecting surface. Using the described procedure, the Time-of-Flight sensor is able to operate in real time at about 30 Hz. It creates dense point clouds, however with a limit spatial resolution. As the measurement principle assumes perfectly sinusoidal signals, which are not achievable in reality, the measured distance is subject to noise. It comprises about 1% of the measured distance. Also the measurement principle is biased as a function of object albedo, resulting in poor performance on textured scenes. A prominent example is the distance measurement of a checkerboard, where the black squares seem closer to the camera than the white squares. Additionally, the quality of the measured intensity image is low.

J. Fischer, G. Arbeiter and A. Verl are with the Robotic Systems Group, Fraunhofer IPA, 70569 Stuttgart, Germany {jan.fischer, georg.arbeiter, alexander.verl}@ipa.fhg.de

Stereo vision estimates depth through triangulation on point correspondences across image pairs and the knowledge of the cameras' intrinsic and extrinsic parameters. On most textured scenes, stereo is able to provide high resolution point clouds. However, in the absence of features, the system fails to measure depth. Due to the different viewing angles of the two cameras, stereo vision is also prone to occlusions. Additionally, low frequency distortions often disturb the feature association, leading to false depth measurements. Current state-of-the-art stereo matching algorithms achieve accurate dense depth maps only when using global optimization algorithms, needing up to a minute of computation time. Only local correlation based methods are fast enough for real time applications, at the cost of less accuracy and sparse depth maps.

The target application of the proposed sensor combination algorithm is based on a catch and carry scenario for the service robot Care-O-bot® 3 [1], [2] as shown in figure 1(a). The robot is equipped with a touch screen panel for convenient user interaction and is able to acquire ordered drinks from a designated area. For the scenario execution the robot must be capable of creating a 3D representation of its environment in order to avoid collisions with obstacles and detect the ordered object within the scene.

This paper gives an overview of existing sensor combination techniques in Section II. The proposed algorithm is described by a detailed description of its single steps in Section III. An emperical evaluation of the sensor combination results is given by Section IV. Section V concludes the paper by summarizing the results and outlining limitations and future work.

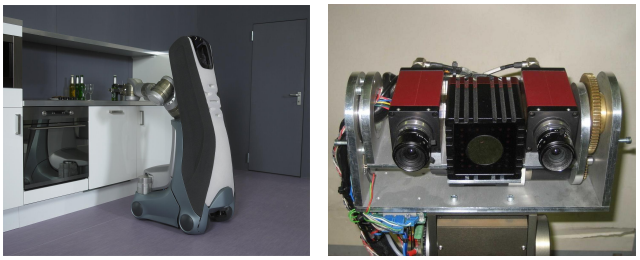
II. RELATED WORK

Several approaches have been proposed to combine the advantages of stereo vision and Time-of-Flight range sensors. In accordance to the proposed taxonomy of Scharstein and Szeliski for stereo algorithms [3], they are roughly categorized in global and local methods. Global methods make explicit smoothness assumptions and solve an optimization problem to minimize a global energy function. Local methods are based on pixelwise matching costs within a certain neighborhood. By selecting the disparity with minimal cost, effectively a winner-take-all optimization is performed. Among others, global methods for sensor combination have been reported by Zhu et al. [4], [5] and Hahne et al. [6]. Zhu et. al. calculate in [4] a depth probability distribution for disparity guesses based on stereo vision and ToF measurements and combine them using Markov Random Fields

(MAP-MRF). In [5] Zhu et al. extend their approach to the temporal domain by using a dynamic Markov Random Field to infer depth from both spatial and temporal neighbors. Like in their previous paper they calculate the maximum-a-posteriori disparity value given the sensor measurements by augmenting the data term of a global optimization algorithm for stereo vision with a term describing the euclidean distance between the corresponding 3D coordinates from the stereo and the Time-of-Flight sensor. Hahne et al. use a graph cut approach in [6], to initialize the domain of the volumetric grid with the depth information from the low resolution Time-of-Flight camera to cut down computation time and increase accuracy of the depth estimation. Local sensor combination approaches have been proposed by Gudmundsson et al. [7], Hahne et al. [8] and Bartczak and Koch [9]. Gudmundsson et al. [7] perform sensor combination by calculating disparity estimates for stereo vision from the 3D Time-of-Flight sensor. This constrains the stereo algorithm on a per pixel basis, resulting in more accurate disparity maps. Hahne et al. combine in [8] a Time-of-Flight sensor and a stereo rig by using data from the Time-of-Flight sensor to limit the disparity search range for the stereo rig using a standard correlation based stereo algorithm. Bartczak and Koch introduce in [9] a cost function for each pixel of a high resolution color image where the minimum of the function's per pixel value corresponds to the optimal depth. The cost function is calculated over local pixel neighborhoods that incorporates the squared distance between the depth measured by the Time-of-Flight sensor and the proposed depth as well as the color consistency of the left and right stereo image patch given the proposed distance. Global methods outperform local approaches in terms of accuracy while, local methods exhibit significantly better performance in terms of computation time. None of the proposed methods elaborates the advantages of stereo systems that reside in between local and global methods to achieve both, acceptable accuracy and computation time. Recently, Hirschmüller [10] proposed a semiglobal matching algorithm for stereo processing. This paper extends the proposed method with the information of the Time-of-Flight sensor to further improve the accuracy and density of the stereo algorithm under acceptable timing conditions.

III. SENSOR COMBINATION

Care-O-bot® 3 is equipped with a multi-sensor camera head consisting of two standard color cameras and one Time-of-Flight camera (Fig. 1(b)). A major advantage of the Time-of-Flight sensor compared to state-of-the-art stereo vision system is its capability to provide depth data for each image pixel in real time. However, state-of-the art stereo systems still significantly outperform current Time-of-Flight sensors in terms of spatial resolution and accuracy. Therefore, the usage of a Time-of-Flight sensor alone is not applicable for many vision tasks and a combination of both systems' advantages is desirable.



(a) Care-O-bot® 3 (b) Sensor head

Fig. 1. Sensor setup of Care-O-bot® 3 for data acquisition. One stereo rig augmented with a Time-of-Flight camera

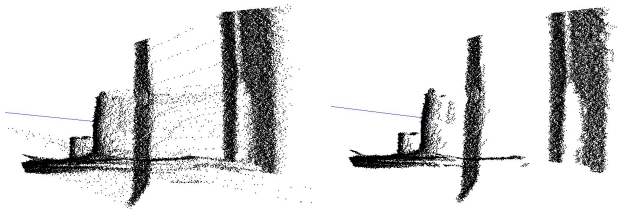
A. Calibration

For intrinsic and extrinsic calibration the method of Zhang [11] using Bouguet's Matlab calibration toolbox [12] has been applied on all cameras. Due to the low resolution and noisy Time-of-Flight intensity data, intrinsic calibration of the Time-of-Flight sensor is prone to be inaccurate. To increase the calibration quality we follow the ideas of Lindner and Kolb [13]. Extrinsic calibration results are used to express 3D data from the Time-of-Flight sensor relative to the coordinate system of the stereo rig, hence enabling an association of measured range from Time-of-Flight data to measured disparity from stereo vision. The stereo rig is initialized using the extrinsic and intrinsic color camera parameters. The resulting colour camera's stereo rectified re-projection matrix enables an association of 3D data with 2D color image coordinates.

B. Pre-processing

For pre-processing, the range values from the Time-of-Flight sensor are filtered to remove speckle from noisy range values. Figure 2(a) shows the raw 3D values delivered by the Time-of-Flight camera of a scene that has artificially be rotated by 90° to show a side-view on the scene. The original color encode range data is shown in Figure 5(b). The most prominent noise originates from tear-off edges on object borders and noise from objects that exceed the non-ambiguity range (0 m to 5 m in case of the Time-of-Flight sensor SwissRanger™ SR4000) which is dependent on the selected modulation frequency. Common filtering techniques apply median filters or fixed amplitude thresholding [8] to remove noisy range measurements. Amplitude filtering removes most of the noise originating from measurements outside the non-ambiguity range. However, neither amplitude filtering nor median filtering effectively remove tear-off edges. We propose to apply wavefront propagation with prior amplitude thresholding for Time-of-Flight data filtering. We iteratively expand the neighborhood of each pixel until a maximal depth threshold t_z, relative to the reference pixel is exceeded. The size s_p of the pixel p's neighborhood is compared against a speckle threshold t_s. When $s_p < t_s$ the pixel's range value is labeled as invalid. Otherwise, it is considered as a valid range value. Using wavefront propagation, only range values of pixels with a sufficient number of close-by (in terms of depth differences) neighbors will survive. This directly

corresponds to the smoothness assumption made for global stereo vision. The results of filtering the 3D Time-of-Flight data using wavefront propagation is shown in Figure 2(b).

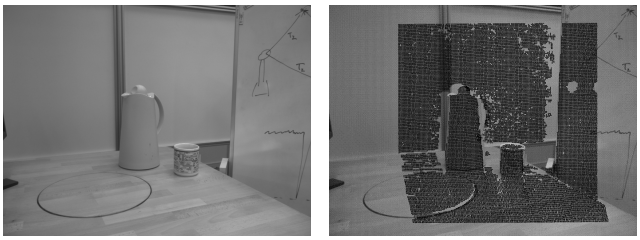


(a) Raw range image (b) Filtered range image

Fig. 2. Filtered and raw range data from Time-of-Flight sensor

C. Re-projection

After filtering, the valid 3D coordinates of the Time-of-Flight camera are re-projected to the image plane of the left color camera 3(b). Let R_t^l be the extrinsic rotation matrix and T_t^l the extrinsic translation vector that relates the left color camera with the Time-of-Flight sensor, then $x_l = R_t^l x_t + T_t^l$ expresses the 3D Time-of-Flight measurement $x_t = (x, y, z)$ relative to the left color camera coordinate system. Using the 4-by-4 stereo rectified re-projection matrix Q_l from the stereo calibration step, the corresponding 2D image coordinates $p_l = (u, v)$ within the left color image and their corresponding disparity d are calculated by $(u', v', d', w)^T = Q_l(x_l, 1)^T$. To retrieve the actual values of p_l and d, u', v' and d' must be normalized by w, $(p, d) = (u'/w, v'/w, d/w)$. The re-projected 3D Time-of-Flight data covers only a small part of the high resolution color image. In order to increase the influence of the re-projected Time-of-Flight measurements on the following semiglobal cost aggregation step (Section III-E) the re-projected Time-of-Flight measurements are propagated to their individual neighborhood using wavefront propagation. To reduce computation time, the propagated range values are not interpolated and copied as they are. To cope with the copied data later on, the uncertainty of each propagated value is increased proportional to the distance from its origin (Section III-D).



(a) Original greyscale image (b) Back-projected 3D Time-of-Flight data (black circles)

Fig. 3. Re-projection of filtered Time-of-Flight data onto the high resolution greyscale image

D. Pixelwise matching costs

The re-projected and propagated Time-of-Flight data serves as an initial disparity guess for each affected pixel of the color image. However, due to occlusion originating from the different viewing angles of color and Time-of-Flight camera, wrong Time-of-Flight range estimates may still have been assigned to several pixels of the color image. To remove these outliers, pixelwise stereo matching costs are calculated on the rectified color image pair. For the calculation of matching costs we apply the pixel dissimilarity measure of Birchfield and Tomasi [14]. It determines the difference of intensities in the range of half a pixel along each direction of the epipolar line. To improve robustness and lower the probability of ambiguous matchings, block matching is applied. The matching costs are accumulated for a squared neighborhood around each pixel resulting in total pixelwise matching costs $C_{BM}(p, d)$ of pixel $p = (u, v)$ and disparity d. Equation 1 states the explained block matching based cost measure

$$C_{BM}(p, d) = \sum_{i=-N}^{N} \sum_{j=-N}^{N} C_{BT}(u - i, v - j) \qquad (1)$$

where $C_{BT}(p)$ corresponds to the dissimilarity measure of Birchfield and Tomasi and N specifies the considered neighborhood.

For each re-projected and propagated Time-of-Flight range measurement, the corresponding pixelwise matching costs $C_{BM}(p, d)$ for the proposed disparity d are compared against a fixed threshold t_{BT}. If $C_{BM}(p, d)$ exceeds t_{BT} it is assumed that the pixel's depth value is incorrect and rejected.

Finally, a discrete cost function is calculated for each pixel. In absence of valid Time-of-Flight measurements, a pixel's cost function represents directly the stereo matching costs calculated to reject the invalid Time-of-Flight measurements and given by Equation 1.

If a valid Time-of-Flight measurement is available for a pixel p, the cost function approximates a reversed Gaussian distribution $1 - P_p$, $P_p = N(\mu, \sigma)$ with μ corresponding to the calculated disparity d based on the Time-of-Flight measurement and σ corresponding to an expected 2% measurement noise relative to the measured distance. The fact that range values from propagated Time-of-Flight data has solely been copied and not interpolated from a close-by pixel is taken into account by increasing the expected measurement noise in proportion to its distance to the measurement's original pixel coordinates. To accelerate computation and to incorporate the possibility that Time-of-Flight based disparity guesses may still be wrong, the cost function becomes constant as disparity differences become larger than 2σ (Equation 2).

$$C_{TOF}(p, d) = \begin{cases} k(1 - P_p(\mu - d)), & \text{if } |d - \mu| < 2\sigma \\ k, & \text{otherwise} \end{cases} \qquad (2)$$

The factor k of Equation 2 corresponds to the maximal costs induced by the cost function $C_{TOF}(p, d)$. The final pixelwise cost function is then given by combining Equation 1 and Equation 2 into Equation 3.

$$C(p,d) = \begin{cases} C_{TOF}(p,d), & \text{if Time-of-Flight data valid} \\ C_{BM}(p,d), & \text{otherwise} \end{cases} \tag{3}$$

E. Cost aggregation

Cost aggregation is based on the work of Hirschmüller [10]. To avoid ambiguous matching costs for different disparities, Hirschmüller connects the pixelwise matching costs with smoothness constraints using the energy function described in Equation 4.

$$\begin{aligned} E(D) = \sum_p (C(p,D_p) + \sum_{q \in N_p} P_1 T[|D_p - D_q| = 1] \\ + \sum_{q \in N_p} P_2 T[|D_p - D_q| > 1] \end{aligned} \tag{4}$$

The function $T(.)$ evaluates to 1 if its argument is true and to 0, otherwise. $C(p, D_p)$ describes the pixelwise matching costs of pixel p at disparity D_p. The remaining terms apply smoothness constraints on the neighborhood of p with the second term penalizing the cost of p with additional costs of P_1, when neighboring pixels have a disparity difference of 1 to p. The third term applies a larger penalty $P_2 > P_1$ for all neighboring pixels, having a disparity difference of larger than 1. Traditionally, energy functions have been optimized using global optimization strategies like simulated annealing [15] or graph cut [16]. However, performing global optimization in 2D is an NP-complete problem and not attractive considering computation time. On the other hand, it is possible to minimize the energy function following individual 1D directions. The 1D optimization can be performed in polynomial time using dynamic programming approaches. However, it possesses the disadvantage of generating streaking effects towards the direction of optimization on the resulting disparity image. In order to avoid the streaking effects, Hirschmüller aggregates matching costs from 16 surrounding 1D directions of each pixel into one cost function equally.

An individual cost path is defined by Equation 5

$$\begin{aligned} L_r(p,d) = & C(p,d) + \min(L_r(p-r,d), \\ & L_r(p-r,d-1) + P_1, \\ & L_r(p-r,d+1) + P_1, \\ & \min_i L_r(p-r,i) + P_2) - \min_k L_r(p-r,k) \end{aligned} \tag{5}$$

where r represents the traversed direction of the individual cost path and i traverses over all disparities except d, $d+1$ and $d-1$. The term $C(p,d)$ corresponds to a pixelwise matching cost function. The first minimization term adds the minimal cost of the preceding pixel for the current optimization direction r, when selecting disparity d for the current pixel. The costs of the preceding pixel are penalized depending on its disparity difference to the currently selected disparity d as explained for Equation 4. In order to keep the value of $L_r(p,d)$ not constantly growing while traversing the cost path, the minimum path cost of the preceding pixel is subtracted from the Equation. The overall costs, including all cost paths are given by Equation 6.

$$S(p,d) = \sum_r L_r(p,d) \tag{6}$$

Selecting the minimum $\min_d S(p,d)$ gives the desired disparity for pixel p.

We adapted the semiglobal optimization approach from Hirschmüller by replacing the proposed pixelwise cost function $C(p,d)$ with the proposed cost function of Equation 3. This enables a combined optimization of disparity costs originating from Time-of-Flight data and stereo data. However, in order to reduce computation time, only the 5 upper individual 1D directions are used as optimization paths.

F. Post-processing

For post-processing the disparity results several post-processing steps according to the paper of Hirschmüller and implementations of standard stereo vision algorithms are performed. At first the uniqueness of the disparity minimizing $S(p,d)$ is tested against all other disparities. The disparity is rejected, if other disparities have similar cost values. Then, a consistency check between the left and right color images based on the selected disparity is performed to ensure that left-right image pixel pairs are unique. To reduce quantization errors originating from the discrete disparity values, the determined disparity d with minimal cost $S(p,d)$ is interpolated over the cost values of the neighboring disparity costs $S(p,d-1)$ and $S(p,d+1)$ using a quadric function 7.

$$d_{new} = d - \frac{S(p,d+1) - S(p,d-1)}{2(S(p,d+1) + S(p,d-1) - 2S(p,d))} \tag{7}$$

Figure 4 gives an overview of the complete procedure.

IV. RESULTS

The proposed algorithm has been tested on a Intel Core 2 Duo CPU with 2.59 GHz. For accelerating computations, the implementation has been optimized using SIMD support to execute basic operations in parallel. The two color cameras have a resolution of 1388x1038 pixels, the SwissRanger™ SR4000 operates with an image resolution of 176x144 pixels. The cost function has been evaluated for 176 different disparities. Figure 5 shows the input images of the two color cameras and the color encoded range data from the Time-of-Flight sensor.

By the use of SIMD based parallelization, the algorithm is able to run on average with 0.3 Hz enabling the procedure to be used for static scenes.

An important factor for the resulting quality of the disparity image is the proportion of re-projected Time-of-Flight range measurements relative to the number of all pixels. On the extremes, having no Time-of-Flight data available will make the algorithm behave like a common

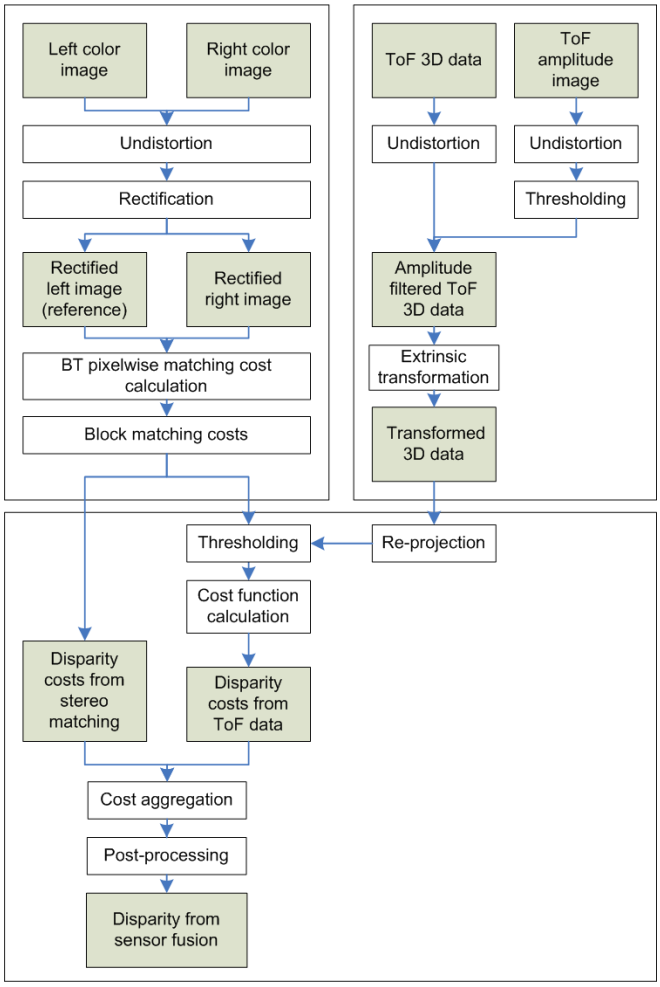
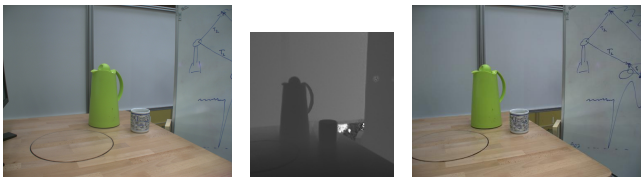


Fig. 4. Schematic overview of sensor combination processing steps

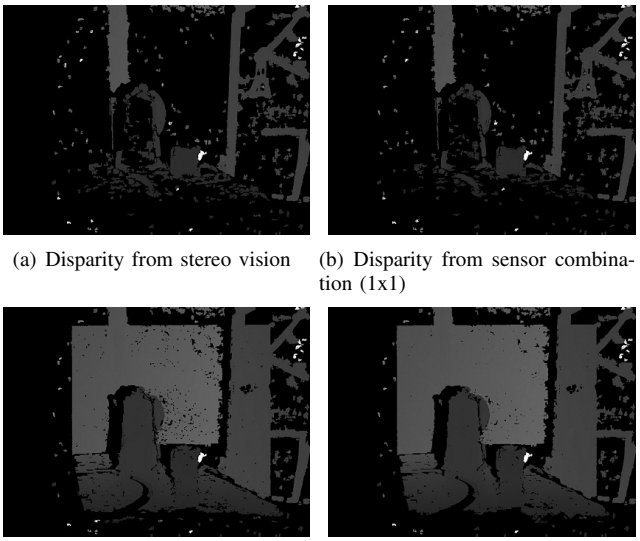


(a) Left color image (b) Time-of-Flight image (c) Right color image

Fig. 5. Input images from all three cameras for sensor combination

stereo algorithm. On the other hand, having Time-of-Flight measurements available for each pixel will make the algorithm to use stereo information solely for filtering wrong Time-of-Flight measurements and the algorithm will apply semiglobal optimization on Time-of-Flight data only. To alter the proportion of available Time-of-Flight data relative to available stereo information, the neighborhood size used for wavefront propagation as described in section III-C may be modified. We compared our algorithm against different ranges for wavefront propagation of the Time-of-Flight data. The results are shown in Figure 6.

The disparity map created without Time-of-Flight information is shown in Figure 6(a). It clearly exhibits the



(a) Disparity from stereo vision (b) Disparity from sensor combination (1x1)

(c) Disparity from sensor combination (3x3) (d) Disparity from sensor combination (5x5)

Fig. 6. Disparity images of stereo vision and sensor combination algorithm

typical drawback of stereo vision with unstructured areas. The white, texture-less walls and the wooden table have no assigned valid disparity values. Figure 6(b) shows the results of relying solely on the re-projected Time-of-Flight measurement without propagating the Time-of-Flight data to its immediate neighboring pixels. The resulting disparity map is quite similar to the disparity map obtained with stereo vision only. This is mainly due to the large number of stereo based disparity estimates compared to the available Time-of-Flight estimates that dominate the semiglobal optimization with their associated cost function. However, the propagation of each pixel's Time-of-Flight measurement to its 3x3 neighborhood results in a significant improvement of disparity density (Figure 6(c)). Now, the re-projected and propagated Time-of-Flight data is in its number almost equal to the number of pixels that rely on stereo based disparity guesses only. Within the optimization procedure of the cost function, the Time-of-Flight values are now able to create unique disparity responses for most of the unstructured areas. The areas where disparity information from Time-of-Flight data is available are clearly visible. The disparity map gets even denser when larger propagation ranges are selected as seen in Figure 6(d). However, increasing the neighborhood range further overwrites effectively the disparity information from stereo vision, which is more accurate in structured areas. It even creates blocky effects due to the propagation of the unmodified original Time-of-Flight values to its neighboring pixels. Therefore, we choose a 5x5 neighborhood to be a suitable propagation size.

The relative amount of pixels with disparity information in proportion to the selected neighborhood size for wavefront propagation of Time-of-Flight measurements and the whole image size is given in Table I. The first column in the table specifies the relative amount of pixels with disparity information using the proposed algorithm without ToF mea-

surements.

0x0	1x1	3x3	5x5
17.46%	16.37%	50.61%	51.77%

TABLE I

DISPARITY DENSITIES IN RELATION TO WAVEFRONT PROPAGATION SIZE OF TIME-OF-FLIGHT MEASUREMENTS

Figure 7 compares the results of stereo vision with the results obtained from the proposed algorithm in combination with Time-of-Flight data.

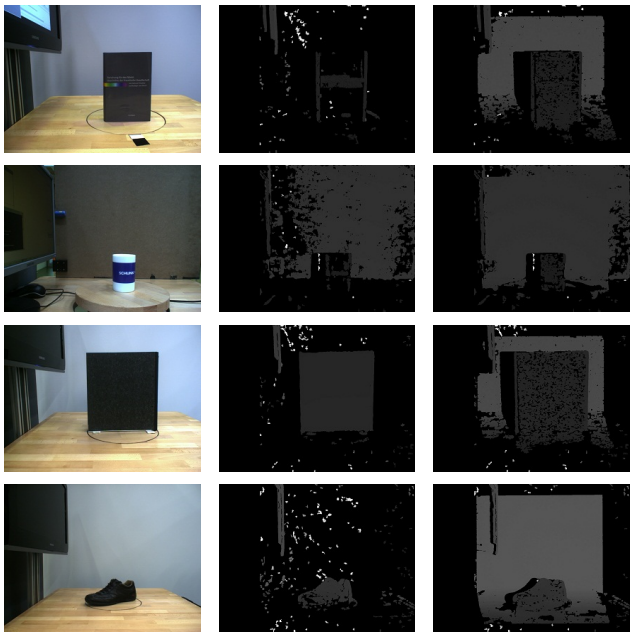


Fig. 7. The first column shows the original color images, the second column disparity images from stereo vision and the third column disparity form the proposed sensor combination algorithm

V. CONCLUSION

In this paper 3D information from different sources has been combined to create denser disparity maps with higher resolution than could be obtained with a single sensor setup alone. The target application of the proposed algorithm is within the context of object detection and pose estimation for service robotics. The proposed algorithm has been implemented on the service robot Care-O-bot® 3 and improves the results from stereo vision especially in structureless areas where stereo vision is prone to errors. The proposed method performs with reasonable computation enabling the investigation of static scenes.

The proposed algorithm does not yet make full usage of the provided disparity estimates from the Time-of-Flight camera. Due to the used dynamic programming approach of the block matching algorithm for pixelwise cost calculation, cost estimates must be calculated for all disparities independent of available disparity estimates from the Time-of-Flight data. Future work will target to decrease computation time by limiting disparity calculation of pixels' with associated Time-of-Flight data to values that are close to its Time-of-Flight estimate. This will reduce computation time to make the method applicable for the 3D recognition of dynamic scenes. Furthermore, accuracy estimates could only visually be conducted as ground truth data is missing. It is the intention to calibrate a tilting laser scanner to the camera system and use its measurements as ground truth for accuracy tests.

VI. ACKNOWLEDGMENTS

The authors gratefully acknowledge that this work was funded as part of the research project "Pool Projekt 2009-22" from the "Centre for Integrative Neuroscience" in Tübingen, Germany.

REFERENCES

[1] C. Parlitz, M. Hägele, P. Klein, J. Seifert, and K. Dautenhahn, "Care-o-bot 3 - rationale for human-robot interaction design," in *Proc. of 39th International Symposium on Robotics (ISR)*, Seoul, Korea, 2008.

[2] U. Reiser, C. Connette, J. Fischer, J. Kubacki, A. Bubeck, F. Weisshardt, T. Jacobs, C. Parlitz, M. Hagele, and A. Verl, "Care-o-bot 3 - creating a product vision for service robot applications by integrating design and technology," in *Intelligent Robots and Systems (IROS 2009)*, oct. 2009, pp. 1992 –1998.

[3] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision*, vol. 47, pp. 7–42, 2002.

[4] J. Zhu, L. Wang, R. Yang, and J. Davis, "Fusion of time-of-flight depth and stereo for high accuracy depth maps," in *CVPR*, 2008, pp. 1–8.

[5] J. Zhu, L. Wang, J. Gao, and R. Yang, "Spatial-temporal fusion for high accuracy depth maps using dynamic MRFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, pp. 899–909, 2009.

[6] U. Hahne and M. Alexa, "Combining time-of-flight depth and stereo images without accurate extrinsic calibration," *Int. J. Intell. Syst. Technol. Appl.*, vol. 5, no. 3/4, pp. 325–333, 2008.

[7] S. A. Gudmundsson, H. Aanaes, and R. Larsen, "Fusion of stereo vision and time-of-flight imaging for improved 3d estimation," *Int. J. Intell. Syst. Technol. Appl.*, vol. 5, no. 3/4, pp. 425–433, 2008.

[8] U. Hahne and M. Alexa, "Depth imaging by combining time-of-flight and on-demand stereo," in *Dyn3D '09: Proceedings of the DAGM 2009 Workshop on Dynamic 3D Imaging*. Berlin, Heidelberg: Springer-Verlag, 2009, pp. 70–83.

[9] B. Bartczak and R. Koch, "Dense depth maps from low resolution time-of-flight depth and high resolution color views," in *ISVC '09: Proceedings of the 5th International Symposium on Advances in Visual Computing*. Berlin, Heidelberg: Springer-Verlag, 2009, pp. 228–239.

[10] H. Hirschmüller, "Stereo processing by semiglobal matching and mutual information," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 30, no. 2, pp. 328–341, feb. 2008.

[11] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1330–1334, 1998.

[12] http://www.vision.caltech.edu/bouguetj/calib_doc/.

[13] M. Lindner and A. Kolb, "Lateral and depth calibration of pmd-distance sensors," in *ISVC*. Springer, 2006, pp. 524–533.

[14] S. Birchfield and C. Tomasi, "Depth discontinuities by pixel-to-pixel stereo," *International Journal of Computer Vision*, vol. 35, pp. 1073–1080, 1996.

[15] J. Marroquin, S. Mitter, and T. Poggio, "Probabilistic solution of ill-posed problems in computational vision," *ASAJ*, vol. 82, no. 397, pp. 76–89, March 1987.

[16] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 11, pp. 1222 –1239, nov. 2001.