Fusion of Stereo Vision and Time-of-Flight Imaging for Improved 3D Estimation

Sigurjón Árni Guðmundsson, Henrik Aanæs and Rasmus Larsen

Technical University of Denmark, Informatics and Mathematical Modelling.
Richard Petersens Plads, Building 321,
2800 Kgs. Lyngby, Denmark
Email: {sag,haa,rl}@imm.dtu.dk

Abstract: In this paper the fusion of two 3D estimation techniques is suggested: Stereo vision and Time-of-Flight (TOF) imaging. By converting the TOF-depth measurements to stereo disparities the correspondence between images from a fast TOF-camera and standard high resolution camera pair are found so the TOF depth measurements can be linked to the image pairs. Also in the same framework a method is developed to initialize and constrain a hierarchical stereo matching algorithm. It is shown that in this way higher spatial resolution is obtained than by only using the TOF camera and higher quality dense stereo disparity maps are the results of this sensor fusion.

Keywords: Time-of-Flight, Stereo Vision, Real-time 3D Computer Vision, Sensor Fusion.

1 Introduction

3D computer vision is a very active research field. Research driven by the need for solutions in such broad fields as robot vision, automatic navigation, computer interaction, tracking and action recognition, only to name a few. Here a method is described that obtains quality 3D estimates using a fusion of the traditional and thoroughly investigated depth from stereo image technique and the new technology of Time-Of-Flight (TOF) range cameras. The main three contributions of this paper are in short:

- A disparity estimate derived from the TOF-depth measurements is found that accurately describes the correspondences between the TOF- and standard camera.
- A method to transfer the TOF-depth to a standard image via the disparity estimate
- Using the disparity estimate to initialize a stereo matching algorithm.

The results show that higher spatial and depth resolution is gained using this fusion than by only using the TOF camera and more correct and higher quality dense stereo disparity maps than using standard stereo methods.

In the remainder of this section the theoretical background is given and some of the related work presented. In section 2 the TOF-stereo rig is introduced and how the TOF-images are fused with the standard images. Section three shows the results of the fusion and the stereo matching algorithm.

1.1 Depth from Stereo

3D from stereo strives at computing corresponding pixels for every pixel of an image pair. The correspondence can be expressed as a disparity vector i.e. if the corresponding pixels are x_l and x_r in the left and right image respectively, then the disparity map $D_l(x_l, x_r)$ is the difference of their image coordinates. The output of a stereo algorithm is therefore this disparity map that maps every pixel from one image to the other. The standard approach is to simplify the disparity vector to one dimension. This is done by performing rectification where the image pairs are transformed by 3×3 homographies, which align epipolar lines to the corresponding scanlines. A detailed description of two view epipolar goemetry can be found in Hartley and Zimmermans book [1].

Most stereo algorithms work by minimizing a cost function and results of various research on such algorithms are reviewed in [2] and are further compared in [3]. The stereo algorithm used here is a very effective, hierarchical, dynamic programming (DP) approach where the matches are made in sequences under order constraints and solved as a global optimization problem. The algorithm is based on Meerbergen et al.'s work [4].

1.2 Depth from Time-of-Flight

The TOF camera used here is a SwissRanger SR3000 [5]. It is designed to be a cost-efficient and eye-safe range image solution.

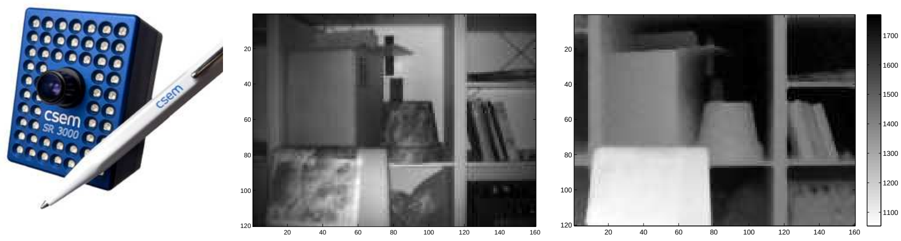

Figure 1 The SwissRanger SR3000 and its output: the intensity image and the range image.

Basically it has an amplitude modulated light source and a two dimensional sensor built in a miniaturized package (see Figure 1).

The sensor is a 176×144 pixel custom designed 0.6 μm CMOS/CCD chip where each pixel in the sensor demodulates the reflected light by a lock-in pixel method

[6]. The sensor reconstructs the returning signal and two images are generated: An intensity (gray scale) image derived from the amplitude of the signal and a range image (depth measurement per pixel) derived from the phase offset of the signal. Figure 1 shows images from the camera.

The SR3000 has it's pros and cons. The cameras low latency makes it very attractive for many applications as it can generate around 20 frames a second and has the depth accuracy of approx. 1 cm at 1 m which is sufficient for various applications [7]. On the other hand the limited spatial resolution of the images and low quality intensity image makes many computer vision techniques difficult to practise. This is why it is essential to develop methods to fuse the information from the TOF camera with other data, e.g. standard high resolution cameras and multiple view depth measurement setups.

1.3 Sensor Fusion

The TOF-sensor technology is a very new technology and thus not much has been done in the field of sensor fusion. PMD technologies [8] is another producer of TOF-sensors with which some experiments in sensor fusion have been done recently. Bedar et al. [9] have done some benchmark comparisons between TOF-measurements and stereo depth measurements using a rig with both a TOF camera and standard cameras. Also Kuhnert et al. [10] experimented with a setup of a stereo rig and tof-camera. Here they took the two measurements and combined them knowing there relative positions and choosing the measurement that gives better results regionally.

The approaches proposed here is very different as the TOF-camera and the standard cameras are fused together through correspondences by assigning the corresponding TOF-measurements to the standard images. Also by using the TOF-measurements to constrain the stereo algorithm per pixel, resulting in a better quality disparity map.

2 The TOF-Stereo Rig

2.1 Design and Calibration

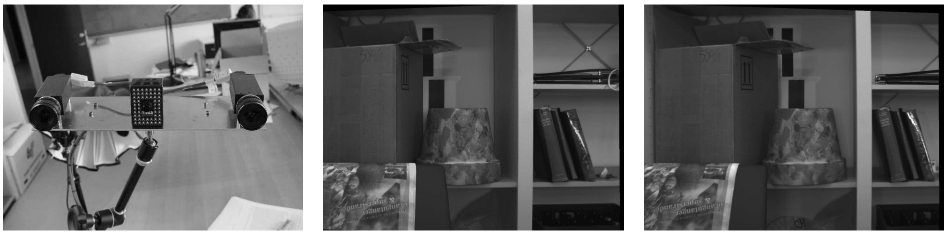

Figure 2 The TOF-Stereo rig and the rectified image pair used in the experiment.

The TOF-Stereo rig shown in Figure 2 is designed to give depth resolution of around 1 mm at 1 m and have a sufficient overlap of the left and right images to be used at 90 cm to 4 m. For this the baseline was chosen 30 cm and the cameras verge towards each other at 7° each from the parallel setup.

All three cameras were calibrated and stereo calibrated using Bouguet's Matlab calibration toolbox [11]. The TOF-images were cropped to 160 × 120 so to have the same format as the standard cameras.

The TOF camera was stereo calibrated with each eight times down-scaled left and right image. The left and right were then calibrated with each other in full size (1280 × 960). The six rectification homographies were found by using Fusiello et al.'s method described in [12].

2.2 Using the TOF-depth to Estimate Disparity

Depth and disparity are interconnected in standard parallel stereo by:

$$Z = \frac{Tf}{D} \tag{1}$$

where Z is the the depth in Euclidian coordinates, T is the baseline, f is the focal length and D is the disparity.

Looking at the rectified left-TOF camera frame, the TOF-depth can be converted to the disparity with the TOF-camera as reference using the relevant baseline and focal lengths getting:

$$D_t(t,l) = \frac{T_{lt} * f}{H^t_{lt} Z} \tag{2}$$

where H^t_{lt} is the homography to rectify the TOF image in the left-TOF pairing. This disparity estimate can now be used to assign the TOF-depth directly to the left image. Results of this merging of TOF and standard images are presented in section 3.1.

Using the TOF-images in the left-right image frame is more complicated. The method used here is based on a method named correspondence linking [13]. The corresponding left and right points have to be linked through the TOF-image, i.e. for each pixel x_t in the TOF-camera image the corresponding pixels x_l and x_r in the other images are found by:

$$\begin{aligned} x_r &= (H^r_{tr})^{-1} D_t(t,r) H^t_{tr} x_t \\ x_l &= (H^l_{lt})^{-1} D_t(t,l) H^t_{lt} x_t \end{aligned}$$

Where H^r_{tr} and H^t_{tr} are the homographies to rectify the right and TOF-image respectively in the right-TOF reference frame and $D_t(t,r)$ is the function that maps the disparity in this frame. The second line does the same in the left-TOF frame.

Calculating x_r and x_l for each point x_t in the TOF-image and transforming them to the left-right rectified frame gives the one dimensional disparity between the left and right images:

$$D_t(l,r) = H^l_{lr} x_l - H^r_{lr} x_r \tag{3}$$

This result is still in the TOF-image's reference frame and needs to be resorted by the left image coordinates resulting in a disparity map estimate for the left reference frame $D_l(l,r)$. Figure 3 shows the results and how the $D_l(l,r)$ 'explodes' when resorted, i.e. there are points in the standard images that have no corresponding points in the TOF image. This is caused by the difference in focal length of the two cameras; they have been scaled to 160 × 120 but the lenses are very different. The grid of no values are filled up by interpolation.

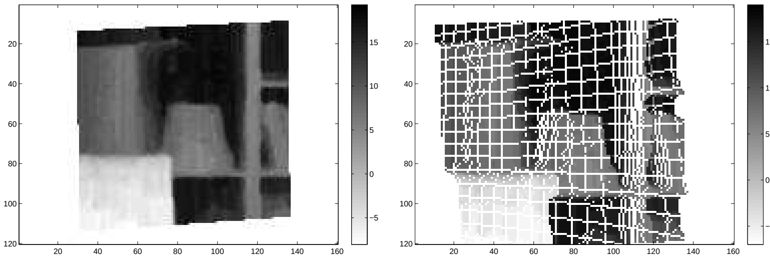

Figure 3 The TOF-disparity estimate $D_l(l, r)$. First in the TOF-cameras reference frame and then rearranged for the left camera. As the focal length of the standard cameras is larger than the TOF-cameras the data "explodes". These holes were filled by interpolation seen in Figure 5.

This is the resulting TOF-disparity estimate hereafter notified by $\hat{D}_l(l, r)$. It is used as a input to the the stereo matching algorithm at the initial 4th level of the hierarchy, i.e. it is used as a 'offset' disparity map to the next higher level. This way the algorithm has a per pixel constraint on the disparity search space.

3 Results

3.1 Fusing TOF Measurements with Standard Image

Figure 4 shows the results of linking the TOF depth measurements to the high resolution image. The results have artifacts directly from the TOF-cameras i.e. the quite low depth resolution. E.g. the edge on the flower pot can barely be detected. The rounding of the estimated disparity points give some erratic behavior close to depth discontinuities but the over all results are not bad.

When the depth map is up-scaled and assigned to the full-size image there are more problems at edges and it still has the artifacts typical to the TOF-depth measurements. This method still gives a good fast but somewhat crude depth estimate, on a high resolution image which can suit well for many applications.

3.2 Hierarchical Stereo Matching using Dynamic Programming

Figure 5 illustrates the results of the disparity calculations on different levels of the hierarchy; both when using $\hat{D}_l(l, r)$ as input and for comparison the standard method of initializing with a disparity range, here given by the range of$\hat{D}_l(l, r)$. The results using $\hat{D}_l(l, r)$ are much better giving correct disparities at difficult areas such as the uniform back wall where the standard approach fails totally. No improvements are made if the standard method is run with a higher level pyramid, the results are the same. Figure 6 shows a cut out of the reconstructed geometry of the flower pot. The edge is very clear and it gives a much smoother and more correct result than from the simple fusion results in section 3.1.

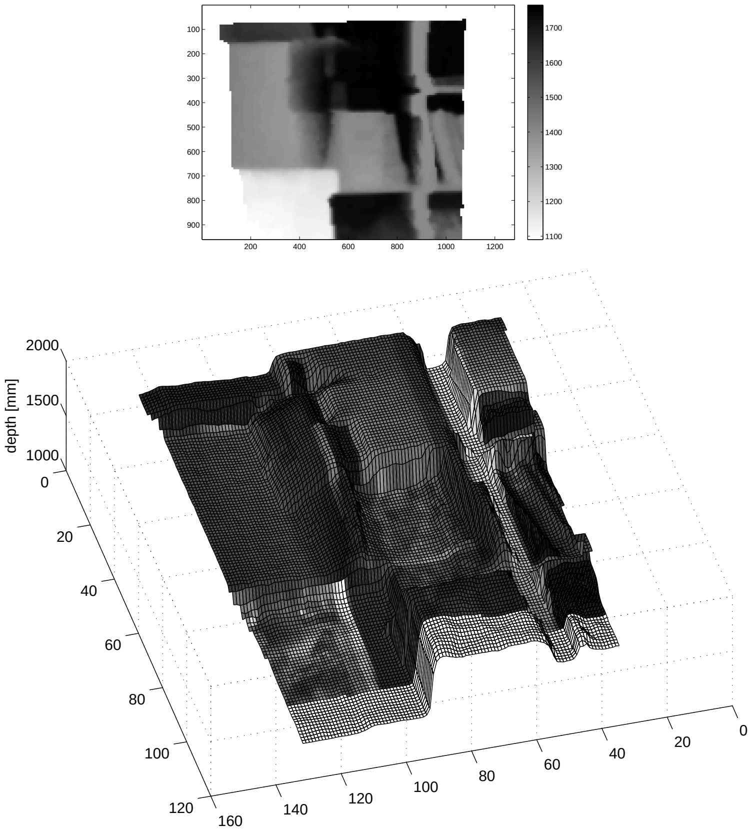

Figure 4 The results of the TOF-left image fusion. Top: The depth map for the Left image. Bottom: Geometric reconstruction in low resolution.

4 Conclusion

A algorithm has been presented to fuse TOF images with images from standard camera pairs in a calibrated framework. The results are promising especially with the TOF-constrained stereo matching algorithm showing much higher quality disparity maps than from the standard approach.

Acknowledgement

This work was in part financed by the Graduate School for Biomedical Optics and New Laser System under a grant from the Danish Ministry of Science, Technology and Innovation. Also by the EU framework 6 IST project ARTTS.

References and Notes

1 R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision.* Cambridge

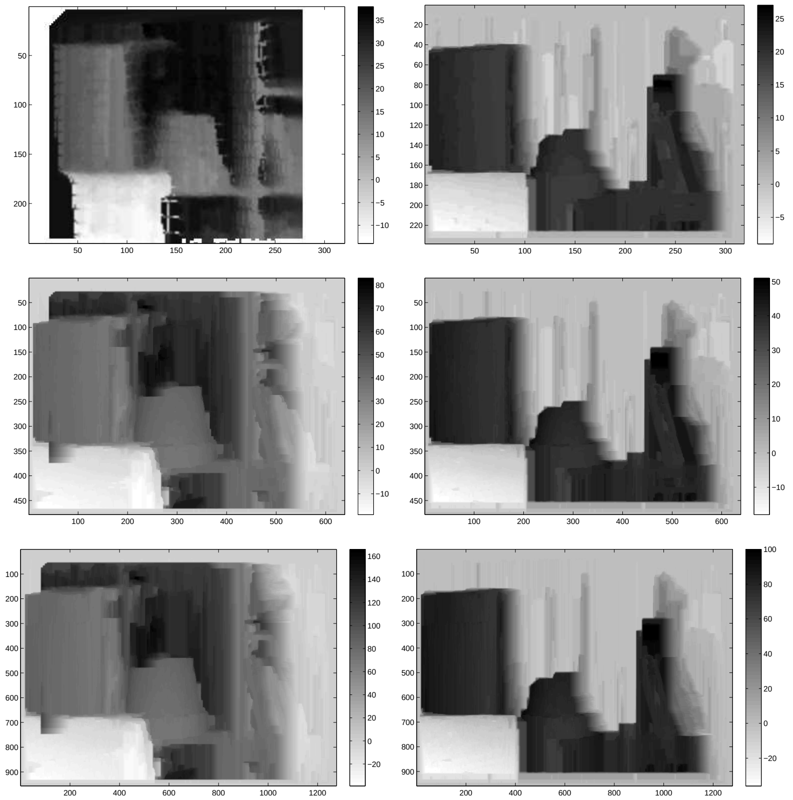

Figure 5 Results from different levels of the hierarchical DP matching. Left column: initialized with $\hat{D}_l(l,r)$ (the first image), then the higher levels of the pyramid Right column: Standard DP method initialized with the range of $\hat{D}_l(l,r)$. The results to the left are much superior. The results to the right totally fail at matching correctly the uniform areas such as the wall in the back.

University Press, ISBN: 0521540518, second edition, 2004.

2 U.R. Dhond and J.K. Aggarwal. Structure from stereo-a review. *IEEE Transactions on Systems, Man and Cybernetics*, 19(6):1489–1510, 1989.

3 D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(3):7–42, 2002.

4 G. Van Meerbergen, M. Vergauwen, M. Pollefeys and L. Van Gool. A hierarchical symmetric stereo algorithm using dynamic programming. *International Journal on Computer Vision*, 47(3):275–285, 2002.

5 http://www.swissranger.ch.

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

7 T. Oggier, B. Büttgen and F. Lustenberger. Swissranger sr3000 and first experiences based on miniaturized 3d-tof cameras. Technical report, CSEM, Swiss Center for Electronics and Microtechnology.

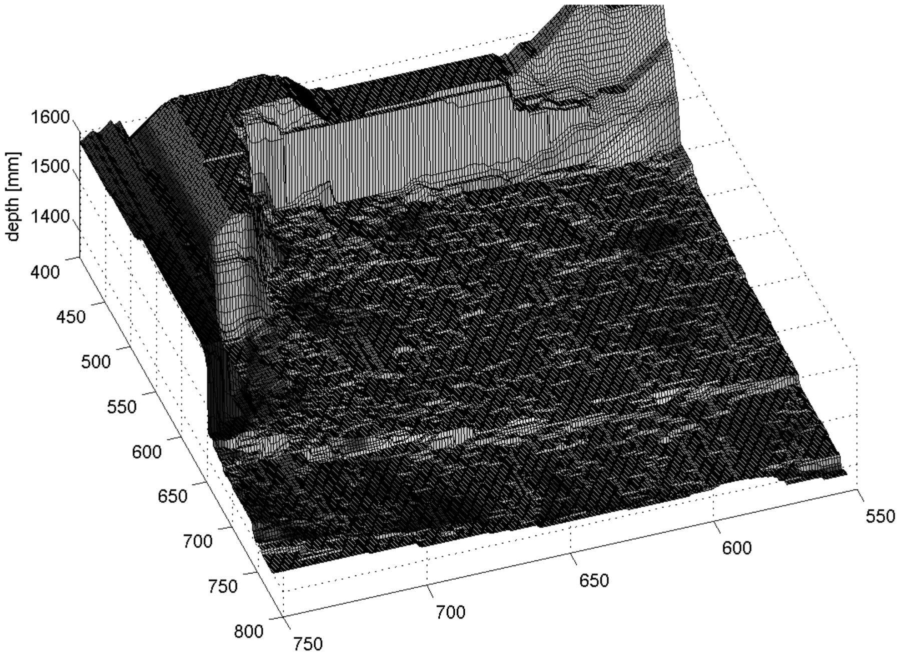

Figure 6 The reconstructed geometry of the flower pot from the final disparity map using the $\hat{D}_l(l,r)$ estimate.

8 http://www.pmdtec.de.

9 C. Beder, B. Bartczak and R. Koch. A comparison of pmd-cameras and stereo-vision for the task of surface reconstruction using patchlets. *The second international ISPRS workshop,BenCOS 2007*, 2007.

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

11 http://www.vision.caltech.edu/bouguetj/calib_doc/.

12 A. Fusiello, E. Trucco and A. Verri. A compact algorithm for rectification of stereo pairs. *Machine Vision and Applications*, 12(1):16–22, 2000.

13 M. Pollefeys, L. Van Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops, and R. Koch Visual modeling with a hand-held camera. *International Journal of Computer Vision*, 59(3):207–232, 2004.