Dense Depth Maps from Low Resolution Time-of-Flight Depth and High Resolution Color Views

Bogumil Bartczak and Reinhard Koch

Institute of Computer Science
University of Kiel
Hermann-Rodewald-Str.3
24118 Kiel, Germany
{bartczak,rk}@mip.informatik.uni-kiel.de

Abstract. In this paper a systematic approach to the processing and combination of high resolution color images and low resolution time-of-flight depth maps is described. The purpose is the calculation of a dense depth map for one of the high resolution color images. Special attention is payed to the different nature of the input data and their large difference in resolution. This way the low resolution time-of-flight measurements are exploited without sacrificing the high resolution observations in the color data.

1 Introduction

The reconstruction of dense depth information for a color image has its use in many applications. Machines or vehicles equipped with color cameras can use the depth information for navigation and map building. A particular advantage of the direct association of depth and color is found in the domain of building digital textured models of scenes and objects. These models serve the purposes of documentation, planing, and visualization. The representation of large scenes and fine details in such models is difficult. Approaches trying to alleviate this challenge are image based rendering techniques [1], where new viewpoints of a scene are generated from a set of color images. These techniques largely benefit from the use of geometric priors, which can have the form of dense depth maps [2]. Due to its compact representation, the image based rendering approach is also attractive to the application in free viewpoint video and 3D television. In order to be useful in future television applications, a dense depth reconstruction scheme has to be able to cope with a large variety of scenes. Unfortunately, the often applied image based matching techniques strongly depend on the scene's texturing. Therefore these approaches tend to fail in homogeneous image regions and at repetitive patterns. An alternative to image matching, is to use active devices like laser scanners or correlating time-of-flight (ToF) cameras [3]. Both devices use the principle of measuring the traveling time of actively sent out light and do not depend on the scene's color distribution. Laser scanners produce high quality scans, but due to their processing scheme and their bulk they

G. Bebis et al. (Eds.): ISVC 2009, Part II, LNCS 5876, pp. 228–239, 2009.

Fig. 1. High resolution color image and corresponding low resolution ToF depth map. Due to the low resolution of the ToF camera the lamp arm is not observed in the depth map although visible in the color image.

are best suited for the reconstruction of static scenes. On the other hand, ToF cameras deliver depth images with high frame rates and are thus applicable in dynamic scenes. ToF cameras however have a low resolution (204×204 pixel)[1] when compared with today's Full HD (1920×1080 pixel) television standard. When combining high resolution color with low resolution ToF depth maps, the mismatch in resolution leads to misalignments at object boundaries. In addition small or thin objects might not be observed by the ToF camera (see fig. 1).

Given multiple high resolution calibrated color views and one calibrated low resolution ToF depth map, this work describes an algorithm for the retrieval of a dense depth map for one of the color images. The combination of color images and ToF depth for the estimation of dense depth maps is not new. However, to our understanding, previous contributions to this specific topic do not regard the different natures of the two data sources properly. This is discussed in the next section. Section 3 then gives an introduction to the proposed approach, while sections 4 and 5 describe the system's parts in detail. In section 6 some results achieved with our algorithm are presented and discussed. The work is concluded in section 7.

2 State of the Art

Different propositions have been made of how to combine ToF data with color information. A rough classification of methods is found by separating approaches that only use a single color view and those that use multiple color views. In [4,5] monocular schemes are discussed. These approaches basically upsample low resolution depth maps by the use of a color controlled bilateral filtering scheme, so that color edges and depth discontinuities become aligned. Although outliers can be detected from a local neighborhood, these algorithms solely rely on the low resolution depth measurements and cannot correctly asses the depth's validity. It is moreover not possible to reintroduce information that has been lost due to the low resolution of the depth map.

If multiple color views are available, the photo consistency across views can be used to estimate depth by triangulation. Dense stereo matching [6] has difficulties with ambiguities, in particular in structureless or repetitive image regions. These

[1] www.pmdtec.com

difficulties can be alleviated in combination with ToF measurements, which are not influenced by such factors. In [4,7,8] approaches are described that initialize a reconstruction with the ToF camera observations and try to refine it using photo consistency. The problem these approaches have to face is that photo consistency is not a convex measure. As such it will only yield improved results with close to correct observations from the ToF camera, and therefore situations like shown in fig. 1 will be difficult to handle. [9] describes an approach for dense depth map estimation where photo consistency and ToF depth maps are evaluated simultaneously. This is achieved by the use of a discrete hypothesis cost volume (see [6]), where the costs per pixel and per hypothesis are a weighted sum of a photo consistency and a robust distance measure to the ToF depth input. Additionally a smoothness cost is introduced and an optimized depth map is derived by using belief propagation. However, two important aspects of this approach are not discussed in the paper.

Firstly ToF depth measurements and the photo consistency have different domains. While the distance measure will give differences in some metric unit, the photo consistency will measure differences in the color value range. A proper combination of these measures does not only have to consider the different value ranges but also the ambiguities in the photo consistency. This means that although it can be expected that a correct hypothesis will have low photo consistency costs, it cannot generally be assumed that a wrong hypothesis will have significantly higher costs. In combination with a wrong ToF depth measurement, this can lead to an error in the reconstruction, although photo consistency on its own might have come to the right conclusion. The second aspect is that due to the interpolation process involved in the upsampling, the ToF depth will appear smooth on the resolution level of the color images. Any sort of smoothness constraint will therefore implicitly prefer the low resolution ToF observations. One way to preserve the high resolution depth information despite smoothness constraints, is to strengthen them via the data term. The definition of a data term well fitting these challenges is the main contribution of this paper.

3 General Approach

The main idea behind the proposed scheme is to find a cost function C for a designated high resolution color image, which will be called the reference image from now on. The positions of the function's per pixel minima shall closely correspond to the correct depth per pixel. This function has to be derived from multiple high resolution color images and a low resolution ToF depth map. The previous section discussed the difficulties when depth measurements and photo consistency are combined. In order to remedy this, we in a first instance estimate depth maps, which are solely based on the color information. This is done via dense local image matching for each (suitable) color image pair containing the reference image. Although it is generally agreed that local image matching is not reliable for dense reconstruction tasks, its use in finding matches between salient image regions is accepted. Instead, a global optimization approach could

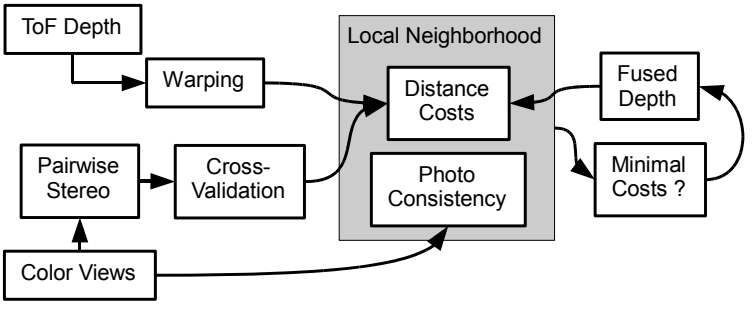

Fig. 2. Overview of the proposed processing scheme

have been applied. At this stage however, it is our interest to reliably find high resolution depth estimates for structures not visible in the ToF depth map. These structures will typically have sufficient saliency, but due to their size, the smoothness constraints used in global optimization might hinder their detection. A scheme well suited for the reconstruction of such structures is proposed in [10]. For the removal of unreliable or wrong matches contained in the estimated depth maps many propositions have been made. Most reliable are the cross-validation of the left/right consistency [11] and to invalidate all depth estimates that lie in image regions where the local color variance is found to be below the anticipated noise level. Other than that, the cost's uniqueness and a maximal cost threshold can be used [12] to find unreliable matches.

After the depth estimation and validation, one or multiple sparse but reliable depth maps for the reference image are available. By using such pairwise stereo estimates rather than simultaneously correlating all views for depth estimation, we follow the idea of multi view stereo reconstruction approaches, where pairwise matches are only used if their costs reach a predefined quality [13,14]. This technique avoids dealing with conflicts in the matching measure, which appear at (semi-)occlusions or non lambertian lighting effects. In such situations, it can be expected that different image pairs will indicate a conflict through inconsistent depth estimates. On the other hand, reliable reconstructions will be indicated through a consensus in the depth estimates. This includes the observations made by the ToF camera.

In order to compare the ToF observations with the high resolution stereo depth maps, we warp the ToF depth map into the reference view. Since this has to be done via a forward mapping from a low to a high resolution, a triangle surface mesh is generated from the ToF depth map and the target map is rendered using computer graphics techniques [15]. Due to the change in viewpoint, wrong occlusions can appear at object boundaries. These disocclusions can be detected by comparing the angle between a triangle's normal and it's line of sight. A triangle is removed if this angle is close to 90 degrees [16].

After these processing steps, a varying number of depth observations is available for each pixel of the reference view. As stated before, we assume that a

consensus in these observations is a good measure for the correctness of a reconstruction. Therefore, the measurement of the consensus is the key motivation when defining the cost functions. Thereby we not only look at the depth observed per pixel but also in a local neighborhood. Since this is a smoothness assumption, which prefers the ToF camera's low resolution, a high resolution photo consistency measure is added as a soft constraint to the cost function. In order to spread the measured consensus, the depth with the lowest costs is selected and fed back into the process. Fig. 2 gives an overview of this process. The details are explained in the next section.

4 Measuring Consensus

In this section, we define the calculation of the cost $C(u, v, z)$ for a depth hypothesis z per pixel (u, v) of the reference image. Like sketched in fig. 2 this cost consists of the distance from the given depth observations and the photo consistency in a local neighborhood. At this point, it has to be considered that the depth distances and the photo consistency have different domains. In order to combine them, we individually transform them into the value range $[0, 1] \subset \mathbb{R}$. Herein, a value of 0 shall indicate that an observation beliefs a hypothesis z to be correct, while the value 1 expresses an observation's disbelief to the correctness of a hypothesis z. Thus, we do not combine values from different domains but rather compare indications to the correctness of a hypothesis. The normalization of all used costs to the range $[0, 1]$ moreover allows to deal with a not fixed number of depth values per pixel. The strategy followed during normalization is to prefer depth hypotheses with the higher count of approving observations.

4.1 Distance Costs

The distance costs $D_i(u, v, z)$ for a hypothesis z per given depth map Z_i are defined as:

$$D_i(u,v,z) = \begin{cases} \min\left((Z_i(u,v) - z)^2 / \mu_Z^2,\, 1 \right) & , Z_i(u,v) \in [z_{\min}, z_{\max}] \\ 1 & , \text{else} \end{cases} . \quad (1)$$

These costs equal 1 if the value $Z_i(u, v)$ found in the depth map is not within the valid search range $[z_{\min}, z_{\max}]$ or if z lies outside the uncertainty range μ_Z. Otherwise D_i is the normalized squared distance between z and $Z_i(u, v)$.

The truncation to the maximal value 1 is not only done to account for outliers, but also to implicitly detect close depth observations. This is visualized in fig. 3. Here cost graphs induced by four hypothetical depth measurements and their mean are depicted. Due to the constant cost contribution outside the uncertainty range, a minimum is formed by the mean, where observations are close to each other. The constant cost contribution, when no valid depth observation is given, allows to implicitly count the number of contributions to a minimum, since costs are potentially lower if more observations are available.

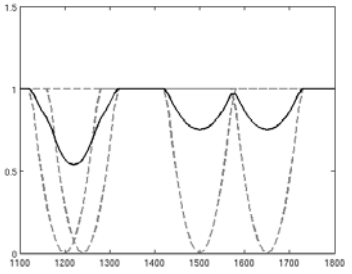

Fig. 3. Hypothetical depth observations and corresponding distance costs (dashed). The consensus on the correctness of a depth is indicated by the strength of a minimum in the mean costs (solid). The maxima in the mean costs reflect the overall agreement on the falseness of a depth.

4.2 Photo Consistency Costs

In order to find an appropriate mapping of the photo consistency into the value range $[0, 1]$, the minimal photo consistency costs induced by a depth hypothesis between the reference image I_r and any other input image I_i are considered:

$$\mathrm{AD}(u, v, z) = \min_i \left(||I_r\left(u, v\right) - I_i\left(H_i\left(u, v, z\right)\right)||_1, \mu_{\mathrm{PC}} \right). \tag{2}$$

In this formula the photo consistency is derived from the absolute color difference observed at hypothetical correspondences. These correspondences are found by the homography mapping H_i utilizing a plane parallel to the reference image plane and with distance z to the reference camera center. The maximal distance value returned is μ_{PC}. This empirically determined threshold expresses the global uncertainty range of the photo consistency measure. It is assumed that the smallest photo consistency costs observed at a pixel corresponds with the correct depth. On the other hand, the largest observed costs will correspond to a wrong depth. Therefore, the final costs PC are defined as follows:

$$\underline{\mathrm{AD}}(u, v) = \min_z \left(\mathrm{AD}(u, v, z) \right), \quad \overline{\mathrm{AD}}(u, v) = \max_z \left(\mathrm{AD}(u, v, z) \right), \tag{3}$$

$$PC(u, v, z) = \begin{cases} \frac{\mathrm{AD}(u,v,z) - \underline{\mathrm{AD}}(u,v)}{\overline{\mathrm{AD}}(u,v) - \underline{\mathrm{AD}}(u,v)} & , \overline{\mathrm{AD}}(u, v) - \underline{\mathrm{AD}}(u, v) > \Delta_{\mathrm{noise}} \\ 1 & , \text{else} \end{cases}. \tag{4}$$

This form of mapping effectively increases the contrast between the lowest and highest observed color difference, which also increases the sensitivity to noise in homogeneous image regions. By comparing the distance between the minimal and maximal values, occlusions and homogeneous image regions can be detected. Therefore we disable the contribution of the photo consistency based costs if the distance range is below the anticipated noise level.

4.3 Aggregation

In the aggregation the distance costs D_i and the photo consistency costs PC observed in a local image neighborhood are fused into a single cost value. This is done in two stages. First, the mean costs per pixel are calculated. Afterwards a weighted mean is applied to combine costs from neighboring pixels. Given N depth maps and a neighborhood $\mathcal{N}(u,v)$ containing (u,v), this is expressed by:

$$C(u,v,z) = \frac{\sum_{(k,l)\in\mathcal{N}(u,v)} w(u,v,k,l,z)\left(\frac{PC(k,l,z)+\sum_{i\leq N} D_i(k,l,z)}{N+1}\right)}{\sum_{(k,l)\in\mathcal{N}(u,v)} w(u,v,k,l,z)}. \quad (5)$$

Like proposed in [10,17] the weights $w(u,v,k,l,z)$ are introduced for the handling of discontinuities. Three features are used to calculate these weights. First it can be expected that with increasing spatial distance of neighboring pixels the observed z distance will differ. Moreover, it can be assumed that depth discontinuities and color discontinuities coincide. So that neighboring pixels with different colors are expected to belong to different depth levels. Finally, jumps in the photo consistency often indicate differences in the depth of neighboring pixels [11]. All three features are expressed in form of distances Δ and transformed into a value range of $[0,1]$ via the function $\omega(\Delta,\gamma) = e^{-\Delta/\gamma}$. The distances used are the spatial distance $\Delta_s = ||(u,v)-(k,l)||_2$, the color distance in the reference image $\Delta_C = ||I_r(u,v)-I_r(k,l)||_2$ and the photo consistency distance $\Delta_{\mathrm{pc}} = |PC(u,v,z)-PC(k,l,z)|$. Using these distances the weights are defined as:

$$w(u,v,k,l,z) = \omega(\Delta_s,\gamma_s)\omega(\Delta_C,\gamma_C)\omega(\Delta_{\mathrm{pc}},\gamma_{\mathrm{pc}}). \quad (6)$$

In the case that $PC(u,v,z)$ is invalid $\omega(\Delta_{\mathrm{pc}},\gamma_{\mathrm{pc}})$ is set to 1.

4.4 Feedback

After aggregation the fused cost for a depth hypothesis z is available, which is based on the number and closeness of observations in a local image neighborhood. Depending on the number of input images and the used approach of acquiring depth observations (section 3) these costs will contain ambiguities, i.e. weak minima. Using a feedback mechanism we try to strengthen minima and at the same time propagate the conclusions drawn from a local neighborhood to adjacent neighborhoods. For this, the depth value with the smallest costs is selected for each pixel, which delivers a fused depth map Z_f. Hereby pixels, whose lowest costs are 1, are assigned the depth value $z_\perp = 0$, which is an invalid depth $z_\perp \notin [z_{\min}, z_{\max}]$ (compare eq. (1)). The fused depth map is added to the set of initial depth maps and the cost functions are recalculated. This can be repeated, replacing Z_f after every iteration, which leads to a diffusion like process. In this process the original input depth maps remain unchanged, which prevents that details become smoothed out. Please observe that due to non linear, adaptive aggregation of costs, discontinuities are still preserved.

5 Refinement

The result of the processing steps described in section 4 is a fused depth map Z_f for a single reference view. Applying this scheme to different views, allows to compare their consistency and this way to detect errors, which in particular will occur at occlusions. Depth values that are not consistent to any other view's fused depth are removed from Z_f by setting them to $z_\perp = 0$. In order to fill the generated holes, we use a simple inpainting technique, where the missing depth values are calculated using the depth values of neighboring pixels with similar color. This is achieved through a monocular adaption of the scheme described in section 4, which thus basically becomes a variant of [4]. For this we only add Z_f into the set of input depth maps and remove the contribution of the photo consistency costs. This change affects the aggregation in eq. (5) and the weight calculation in eq. (6). The repetitive application of this monocular adaption, diffuses the mean depth value of border pixels into invalidated regions, if the border pixels have a similar color like the invalidated pixels. Pixels within larger holes are not filled immediately, since their minimal costs remain 1 until the hole sizes shrink sufficiently (compare section 4.4 and eq. (1)).

6 Experiments

We tested our proposals on two real indoor data sets, consisting of four color views in Full HD resolution (1920×1080 pixels) and one low resolution ToF depth map (176×144 pixels). This data is shown in fig. 4 and 5. We used the approach described in [18] to find the cameras' parameters, thereafter we estimated three sparse stereo depth maps for each color view and warped the ToF depth map into every color view, like described in section 3. Fig. 6 shows an extract of the calculated input depth maps. These input depth maps were fused into four depth maps like discussed in section 4. Finally, a refined depth map was calculated for one of the views, following the explanations in section 5. In order to find a close approximation to the depth value with minimal costs, the cost function was densely sampled using M equally spaced hypotheses in the search range, which in the given situation was app. $z_{\min} = 2m$ to $z_{\max} = 7m$. After selecting the depth sample with the lowest costs $\tilde{z}$ the final depth approximate was found by computing the minimum of a quadratic polynomial, which was fitted to the z positions $\tilde{z} - \Delta_z$, $\tilde{z}$,$\tilde{z} + \Delta_z$ and their corresponding costs. The same sampling was used to determine approximations to the minimal and maximal matching costs $\underline{\mathrm{AD}}$ and $\overline{\mathrm{AD}}$. The parameters used for the fusion and the refinement are listed in table 1. The right columns of fig. 7 show the results we achieved with this approach. The center right column shows the result if only a single camera pair is used. The right most column displays the fused depth maps if all camera pairs are combined. The gain in using multiple pairs is on the one hand the increased number of depth observations for fine structures. On the other hand it allows to resolve occlusion problems. This is clearly visible when comparing the top images of fig. 7. Due to the use of multiple image pairs, the reconstruction

Table 1. Used parameters during the fusion and the refinement

Parameter	Fusion	Refinement
Number of Iterations	5	5
Distance Tolerance	$\mu_Z = 7.5$cm	$\mu_Z = 3.75$cm
Number of Hypotheses (Spacing)	$M = 260$ ($\Delta_z \approx 2$cm)	$M = 260$ ($\Delta_z \approx 2$cm)
Neighborhood	41×41pxl	81×81pxl
Spatial distance weight	$\gamma_S = 9.5$	$\gamma_S = 19$
RGB-Color distance weight	$\gamma_C = 13.1$	$\gamma_C = 4.35$
PC-distance weight	$\gamma_{\mathrm{pc}} = 0.1$	not used
AD-truncation	$\mu_p c = 25$	not used
Expected noise level	$\Delta_{\mathrm{noise}} = 9$	not used

in the area between the leaves of the plant in the upper right corner is improved as well as the image part between the person's chin and its torso.

In order to show the improvements between our approach and previous propositions (see section 2) the depth maps shown in the left columns of fig. 7 were calculated. The left most column displays the outcome of using the monocular scheme proposed in [4]. Due to the strong dependence on the ToF data this approach fails at fine structures. This can be observed at the plants' leaves in the top row or the lamp arm in the lower image row. Moreover erroneous measurements cannot be detected. These results however demonstrate that the use of low resolution ToF depth can be an asset, especially in the structureless regions where color matching is questionable. In the left center column the effect of adding a photo consistency cost like proposed in [9] can be seen. Some severe errors in the top left corner of the ToF depth in the beergarden data set are improved. Furthermore fine structures and edges are reconstructed with greater precision. However the lamp arm in the living room data was not recovered. The reason for this is the ambiguity in the photo consistency, especially when using a window based measure. The soft color adaptive steering of the aggregation, like used in [9], does in some cases not suffice to prevent the bias towards wrong depth hypotheses. The situation is worsened if those wrong hypotheses are additionally supported by the depth measurements of the ToF camera. Three mechanisms in our algorithm address this problem. Firstly the use of cross validated stereo depth maps, which deliver an unbiased high resolution opinion on the depth distribution from the color views. Secondly the normalization of the photo consistency and the distance costs to the range of $[0, 1]$. This way the pixel based photo consistency can compete with the distance costs. Finally the integration of the photo consistency distance into the adaptive aggregation (see eq. (6)), so that discontinuities are handled properly even if the required color contrast is too low.

Please observe that despite the feedback mechanism and the refinement (section 5) the reconstruction is derived from local observations alone. Nevertheless the achieved results (fig. 7) are very homogeneous. In many cases this is owed to the low resolution ToF depth maps. Due to the increased contribution of the high resolution color data in our proposal, the results achieved with only two

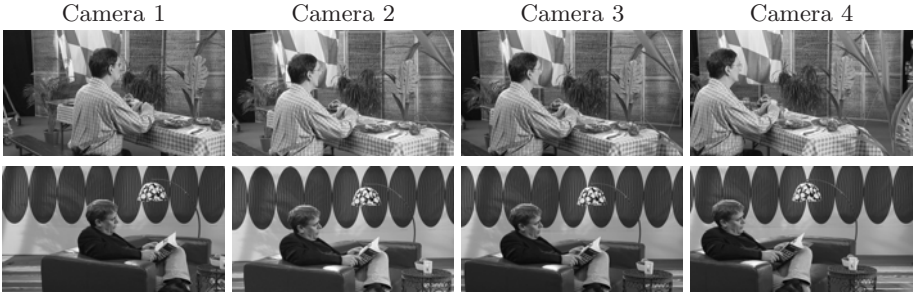

Fig. 4. Captured color views of the beergarden scene (top row) and the living room scene (bottom row)

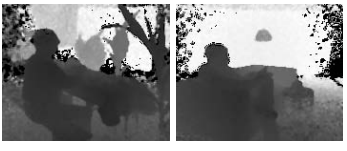

Fig. 5. Captured depth maps: beergarden scene (left) and living room scene (right)

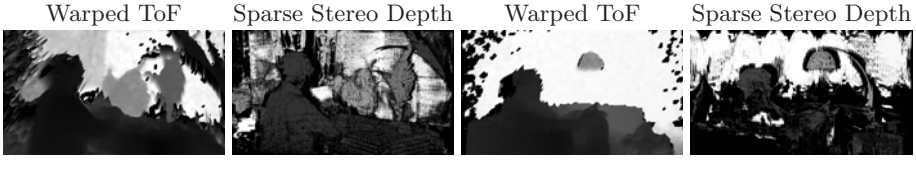

Fig. 6. Excerpt of used input depth maps for camera 3. The stereo depth maps shown are generated from matches between cameras 2 and 3.

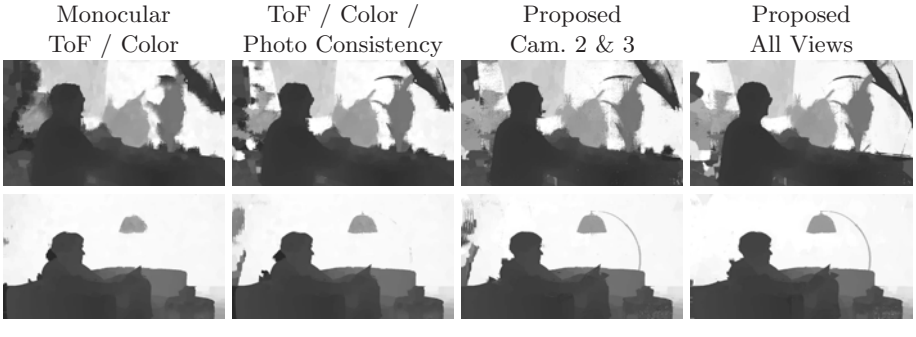

Fig. 7. Results achieved for camera 3. Left: color controlled upsampling of ToF depth maps. Center Left: simultaneous evaluation of distance costs and photo consistency. Center Right and Right: proposed approach using camera views 2 and 3 or using all views.

color views, can become slightly more noisy than the results of previous works. On the contrary it was shown that the proposed cost term allows to reliably preserve high resolution structures, which makes it a good candidate for a data term in approaches that integrate smoothness models [6].

7 Conclusion

In this work we discussed a robust approach to the fusion of low resolution ToF depth images and stereo reconstruction from high resolution color images. The presented algorithm can exploit the use of more than two color views and delivers dense depth maps for a high resolution color image. Fine structures are well preserved even if contradictions are present in the low resolution ToF depth maps. In the current state the approach is heuristic in nature. However the use of uncertainty ranges and normalization already brings it close to a probabilistic formulation. The investigation and proper modeling of uncertainties will be the concern of future work, especially in the face of the additional information that are available from ToF cameras. Future work will also deal with the integration of sophisticated smoothness constraints to resolve ambiguities. Like described in section 2 the challenge here is to avoid a bias towards the low resolution ToF depth maps in the reconstruction.

Acknowledgment

This work was partially supported by the German Research Foundation (DFG), KO-2044/3-2 and the Project 3D4YOU, Grant 215075 of the Information Society Technologies area of the EUs 7^{th} Framework programme.

References

1. Buehler, C., Bosse, M., McMillan, L., Gortler, S., Cohen, M.: Unstructured lumigraph rendering. In: SIGGRAPH 2001: Proceedings of the 28th annual conference on Computer graphics and interactive techniques, pp. 425–432. ACM, New York (2001)
2. Shade, J., Gortler, S., He, L.w., Szeliski, R.: Layered depth images. In: SIGGRAPH 1998: Proceedings of the 25th annual conference on Computer graphics and interactive techniques, ACM, New York (1998)
3. Xu, Z., Schwarte, R., Heinol, H., Buxbaum, B., Ringbeck., T.: Smart pixel - photonic mixer device (pmd). In: M2VIP 1998 - International Conference on Mechatronics and Machine Vision in Practice, pp. 259–264 (1998)
4. Yang, Q., Yang, R., Davis, J., Nister, D.: Spatial-depth super resolution for range images. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition CVPR 2007, pp. 1–8 (2007)
5. Chan, D., Buisman, H., Theobalt, C., Thrun, S.: A noise-aware filter for real-time depth upsampling. In: Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications with ECCV (2008)

6. Scharstein, D., Szeliski, R., Zabih, R.: A Taxonomy and Evaluation of Dense Two-frame Stereo Correspondence Algorithms. In: Proceedings of IEEE Workshop on Stereo and Multi-BaselineVision, Kauai, HI (2001)
7. Beder, C., Bartczak, B., Koch, R.: A combined approach for estimating patchlets from PMD depth images and stereo intensity images. In: Hamprecht, F.A., Schnörr, C., Jähne, B. (eds.) DAGM 2007. LNCS, vol. 4713, pp. 11–20. Springer, Heidelberg (2007)
8. Gudmundsson, S., Aanaes, H., Larsen, R.: Fusion of stereo vision and time-of-flight imaging for improved 3d estimation. In: Proceedings of the DAGM Dyn3D Workshop, Heidelberg, Germany (2007)
9. Zhu, J., Wang, L., Yang, R., Davis, J.: Fusion of time-of-flight depth and stereo for high accuracy depth maps. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition CVPR 2008, pp. 1–8 (2008)
10. Yoon, K.-J., Kweon, I.S.: Adaptive support-weight approach for correspondence search. IEEE Transactions of Pattern Analysis and Machine Inteligence (PAMI) 28(4), 650–656 (2006)
11. Egnal, G., Egnal, G., Wildes, R.: Detecting binocular half-occlusions: empirical comparisons of five approaches. IEEE Transactions of Pattern Analysis and Machine Inteligence (PAMI) 24, 1127–1133 (2002)
12. Muhlmann, K., Maier, D., Hesser, R., Manner, R.: Calculating dense disparity maps from color stereo images, an efficient implementation. In: Proc. IEEE Workshop on Stereo and Multi-Baseline Vision (SMBV 2001), pp. 30–36 (2001)
13. Goesele, M., Curless, B., Seitz, S.: Multi-view stereo revisited. In: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 2402–2409 (2006)
14. Furukawa, Y., Ponce, J.: Accurate, dense, and robust multi-view stereopsis. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition CVPR 2007, pp. 1–8 (2007)
15. Bartczak, B., Schiller, I., Beder, C., Koch, R.: Integration of a time-of-flight camera into a mixed reality system for handling dynamic scenes, moving viewpoints and occlusions in real-time. In: Proceedings of the 3DPVT Workshop, Atlanta, GA, USA (2008)
16. Pajarola, R., Sainz, M., Meng, Y.: Dmesh: Fast depth-image meshing and warping. Int. J. Image Graphics 4, 653–681 (2004)
17. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: Proc. Sixth International Conference on Computer Vision, pp. 839–846 (1998)
18. Schiller, I., Beder, C., Koch, R.: Calibration of a pmd camera using a planar calibration object together with a multi-camera setup. In: Proceedings of the ISPRS Congress, Bejing, China (2008)