A New Multi-lateral Filter for Real-Time Depth Enhancement

Frederic Garcia[⋆†*] Djamila Aouada[⋆] Bruno Mirbach[†] Thomas Solignac[†] Björn Ottersten[⋆]
[†]Advanced Engineering - IEE S.A. [⋆]SnT - Universtity of Luxembourg
{frederic.garcia, bruno.mirbach, thomas.solignac}@iee.lu {djamila.aouada, bjorn.ottersten}@uni.lu

Abstract

We present an adaptive multi-lateral filter for real-time low-resolution depth map enhancement. Despite the great advantages of Time-of-Flight cameras in 3-D sensing, there are two main drawbacks that restricts their use in a wide range of applications; namely, their fairly low spatial resolution, compared to other 3-D sensing systems, and the high noise level within the depth measurements. We therefore propose a new data fusion method based upon a bilateral filter. The proposed filter is an extension the pixel weighted average strategy for depth sensor data fusion. It includes a new factor that allows to adaptively consider 2-D data or 3-D data as guidance information. Consequently, unwanted artefacts such as texture copying get almost entirely eliminated, outperforming alternative depth enhancement filters. In addition, our algorithm can be effectively and efficiently implemented for real-time applications.

1. Introduction

Time-of-Flight cameras are a relatively new 3-D sensing sensors that promise to be an alternative to other 3-D sensing systems such as stereo vision systems, laser scanners or structured light. They present several advantages such as simultaneously providing intensity and distance information for every pixel at a high frame rate. Moreover, they are compact, robust to illumination changes and of low weight. Although ToF cameras cannot yet attain the resolution and precision of alternative 3-D sensing systems, their distinctive features make them suitable for many applications where not very precise but fast 3-D image data acquisition is needed, such as coarse 3-D reconstruction, obstacle avoidance, human tracking and pose estimation among others [4, 8]. Indeed, in applications where the limited resolution of a ToF camera is critical, it is complemented with other sensors, usually colour cameras [5]. For this reason, data fusion is a very promising strategy to overcome ToF drawbacks [1, 3, 6, 7, 9, 17, 19]. Indeed, current research efforts in ToF and 2-D data fusion deliver dense depth maps at near real-time frame rates, outperforming, in some cases, alternative 3-D sensing systems.

In this paper, we propose an adaptive data fusion method for depth enhancement. Indeed, we extend our previous multi-lateral filter based upon a bilateral filter, namely, the pixel weighted average strategy (PWAS), recently proposed in [6]. The PWAS filter copes well with inaccurate edges as it includes a new factor that explicitly accounts for the unreliability of the depth measurement along the edges. Although this filter outperforms alternative data fusion methods for depth enhancement, some limitations are yet to be improved, such as the artefacts due to texture copying. We therefore propose an extension of the PWAS filter and define a new adaptive filter. Our contribution consists in considering not only the 2-D information, but also the depth measurements when filtering depth measurements describing a smooth geometry. This combination will allow to significantly reduce the noise level in depth measurements and to almost entirely eliminate unwanted artefacts. Moreover, in order to ensure that this filter maintains a high computational efficiency for real-time applications, we propose to adopt the fast bilateral filter implementation proposed by Yang et al. [16].

The outline of this paper is as follows: In Section 2, we cover state-of-the-art techniques in low-level data fusion for low-resolution depth map enhancement. In Section 3, we present our new adaptive filter, and propose, in Section 4, its implementation for a real-time data fusion. In Section 5, we compare and quantify our filter with alternative depth enhancement methods. Finally, in Section 6, we give our conclusions.

2. Related Work and Background

When the low-resolution provided by ToF cameras restricts their use in certain applications, the combination with colour cameras is a valid approach to enhance and improve the ToF data. Data fusion exploits the advantages of each of the cameras in the hybrid ToF multi-camera rig avoiding their individual drawbacks. We talk about low-level data

*F. Garcia was supported by the AFR Grant Scheme (Aides à la Formation-Recherche), managed by the National Research Fund Luxembourg (FNR). AFR Grant: TR-PHD BFR08-120.

fusion in contrast to higher fusion levels in which the fusion deals with post-processed data (feature or decision fusion) [11]. Over the last years, there have been some attempts for data fusion. The application of Markov Random Fields (MRFs) to the problem of enhancing ToF data by considering both ToF and 2-D data was first presented by Diebel et al. [3], and extended by Gloud et al. [7]. Despite their promising results, both methods are not suitable if real-time is required. To that end, more recent sensor fusion approaches based upon a bilateral filter, an edge-preserving image filter [14], enables a real-time data fusion. This was achieved by adapting recent implementation techniques that accelerate the bilateral filtering process [12, 13, 15, 16]. Kopf et al. presented the Joint Bilateral Upsampling (JBU) filter [9], a modification of the bilateral filter that considers two different data sources within the kernel of the filter. This way, they upsample the downsampled data for image analysis and enhancement tasks, such as tone mapping, to the full resolution input image. This idea is the basis of our approach as it also applies for depth map enhancement as further investigated by Crabb et al. [2] for real-time matting. The JBU filter has a spatial weighing term $f_{\mathbf{S}}(\cdot)$ based on the pixel position, and a range weighing term $f_{\mathbf{I}}(\cdot)$ based on the 2-D data. Thus, this filter adjusts the edges in the low-resolution depth map $\mathbf{R}$ to the edges in the 2-D guidance image $\mathbf{I}$ as follows:

$$\mathbf{J_1}(\mathbf{p}) = \frac{\sum_{\mathbf{q}\in N(\mathbf{p})} f_{\mathbf{S}}\big(\mathbf{p},\mathbf{q}\big) f_{\mathbf{I}}\big(\mathbf{I}(\mathbf{p}),\mathbf{I}(\mathbf{q})\big)\mathbf{R}(\mathbf{q})}{\sum_{\mathbf{q}\in N(\mathbf{p})} f_{\mathbf{S}}(\mathbf{p},\mathbf{q}) f_{\mathbf{I}}\big(\mathbf{I}(\mathbf{p}),\mathbf{I}(\mathbf{q})\big)}, \quad (1)$$

where $N(\mathbf{p})$ is the neighbourhood at the pixel indexed by the vector $\mathbf{p} = (i,j)^T$, with i and j indicating the row, respectively column of the pixel. The weighing terms $f_{\mathbf{S}}(\cdot)$ and $f_{\mathbf{I}}(\cdot)$ are generally taken to be Gaussian functions with variances $\sigma_{\mathbf{S}}$ and $\sigma_{\mathbf{I}}$, respectively. The resulting filtered image $\mathbf{J_1}$ is an enhanced version of $\mathbf{R}$, that presents less discontinuities, and a reduced noise level. Nevertheless, according to the bilateral filter principle, the fundamental heuristic assumptions about the relationship between depth and intensity data may lead to erroneous copying of texture into actually smooth geometries on the depth map. Furthermore, a second unwanted artefact known as edge blurring appears along depth edges that have no corresponding edge in the 2-D image, *i.e.,* in situations where the objects on either side of a depth discontinuity have similar colour. Chan et al. proposed in [1] the so-called Noise Aware Filter for Depth Upsampling (NAFDU), an extension of the JBU filter that prevents artefacts in those areas where JBU is likely to cause erroneous texture copy. Specifically, the NAFDU appraoch makes the filter behave like the JBU filter except in the areas that are geometrically smooth but heavily contaminated with random noise in the distance measurements. In that case, the filter behaves like a standard bilateral filter, *i.e.,* analysing both spatial and range domains on the same depth map $\mathbf{R}$ without considering the guidance image information. The crux is to decide when the filter has to switch from one case to the other, which for NAFDU needs to be manually tuned.

More recently, we proposed the PWAS filter [6], an alternative extension to the JBU filter that copes well with inaccurate edges as it includes an additional factor $\mathbf{Q}(\cdot)$ to the kernels in (1), named credibility map, and defined as a weighted gradient of the low-resolution depth map $\mathbf{R}$, $\mathbf{Q}(\mathbf{q}) = f_{\mathbf{Q}}(|\nabla\mathbf{R}(\mathbf{q})|)$, and $f_{\mathbf{Q}}(\cdot)$ being a weighing function. Due to the low spatial resolution of ToF cameras, there are pixels that cover foreground and background at the same time. Thus, distance measurements at depth edges may be inaccurate or erroneous. This factor $\mathbf{Q}(\cdot)$ explicitly accounts for these pixels. In a nutshell, the credibility map boundaries define in which areas the distance measurements are unreliable and thus adjusted according to the 2-D guidance image. The PWAS filter takes the form of:

$$\mathbf{J_2}(\mathbf{p}) = \frac{\sum_{\mathbf{q}\in N(\mathbf{p})} f_{\mathbf{S}}(\mathbf{p},\mathbf{q}) f_{\mathbf{I}}\big(\mathbf{I}(\mathbf{p}),\mathbf{I}(\mathbf{q})\big)\mathbf{Q}(\mathbf{q})\mathbf{R}(\mathbf{q})}{\sum_{\mathbf{q}\in N(\mathbf{p})} f_{\mathbf{S}}(\mathbf{p},\mathbf{q}) f_{\mathbf{I}}\big(\mathbf{I}(\mathbf{p}),\mathbf{I}(\mathbf{q})\big)\mathbf{Q}(\mathbf{q})}. \quad (2)$$

Unwanted artefacts such as blurring of depth edges with no corresponding edge in the guidance image may still occur, but are significantly reduced compared to the JBU filter.

3. Proposed depth enhancement filter

Although the PWAS filter outperforms alternative data fusion techniques for depth enhancement as demonstrated in [6], the assumption of only considering the 2-D guidance image within the range term of the kernel in (2) may entail to texture copying in regions that are actually geometrically smooth. Instead, one should directly consider the depth map measurements as, in that situations they usually are reliable. Thus, we define two separate normalized kernels with each one operating on a different data source. The decision of which kernel the filter has to consider is automatically given by the credibility weight of the pixel to be filtered. Therefore, the main benefit of our assumption is the increase of the depth measurement accuracy within smooth regions. The proposed filter takes the form of:

$$\mathbf{J_4}(\mathbf{p}) = \big(1 - \mathbf{Q}(\mathbf{p})\big) \cdot \mathbf{J_2}(\mathbf{p}) + \mathbf{Q}(\mathbf{p}) \cdot \mathbf{J_3}(\mathbf{p}), \quad (3)$$

where $\mathbf{J_3}(\mathbf{p})$ is the filtered value at pixel $\mathbf{p}$ given by a PWAS filtering considering only the depth information from $\mathbf{R}$, *i.e.,*

$$\mathbf{J_3}(\mathbf{p}) = \frac{\sum_{\mathbf{q}\in N(\mathbf{p})} f_{\mathbf{S}}(\mathbf{p},\mathbf{q}) f_{\mathbf{R}}\big(\mathbf{R}(\mathbf{p}),\mathbf{R}(\mathbf{q})\big)\mathbf{Q}(\mathbf{q})\mathbf{R}(\mathbf{q})}{\sum_{\mathbf{q}\in N(\mathbf{p})} f_{\mathbf{S}}(\mathbf{p},\mathbf{q}) f_{\mathbf{R}}\big(\mathbf{R}(\mathbf{p}),\mathbf{R}(\mathbf{q})\big)\mathbf{Q}(\mathbf{q})}. \quad (4)$$

Similarly to [6], we choose the weighting functions to be Gaussian functions with variances $\sigma_{\mathbf{S}}$, $\sigma_{\mathbf{I}}$, $\sigma_{\mathbf{R}}$ and, $\sigma_{\mathbf{Q}}$, respectively. We set the $\sigma_{\mathbf{S}}$ to the scale factor between the

low-resolution depth map and the guidance image. $\sigma_{\mathbf{I}}$ and $\sigma_{\mathbf{R}}$ are set to the mean of the gradient of images $\mathbf{I}$ and $\mathbf{R}$, respectively. The value of $\sigma_{\mathbf{Q}}$ corresponds to the mean of the noise level in the depth map measurements.

4. Implementation

In order to ensure that our filter maintains a high computational efficiency for real-time applications, we adopted the bilateral filter implementation proposed in [16], where Yang et al compared their real-time bilateral filtering against state-of-the-art methods [12, 13] outperforming for accuracy, speed and memory consumption. Thus, we have adapted this method to the previously presented filter by defining four mappings: $E^{\mathbf{I}(\mathbf{p})}(\cdot)$ and $F^{\mathbf{I}(\mathbf{p})}(\cdot)$ for a fixed value of the 2-D image $\mathbf{I}$ at the pixel $\mathbf{p}$ and $G^{\mathbf{R}(\mathbf{p})}(\cdot)$ and $H^{\mathbf{R}(\mathbf{p})}(\cdot)$ for a fixed value of the depth map $\mathbf{R}$ at the pixel $\mathbf{p}$, such that:

$$\begin{aligned} E^{\mathbf{I}(\mathbf{p})}: &\quad \mathbf{q} \longmapsto f_{\mathbf{I}}\big(\mathbf{I}(\mathbf{q}),\mathbf{I}(\mathbf{p})\big)\cdot\mathbf{Q}(\mathbf{q})\cdot\mathbf{R}(\mathbf{q}),\\ F^{\mathbf{I}(\mathbf{p})}: &\quad \mathbf{q} \longmapsto f_{\mathbf{I}}\big(\mathbf{I}(\mathbf{q}),\mathbf{I}(\mathbf{p})\big)\cdot\mathbf{Q}(\mathbf{q}),\\ G^{\mathbf{R}(\mathbf{p})}: &\quad \mathbf{q} \longmapsto f_{\mathbf{R}}\big(\mathbf{R}(\mathbf{q}),\mathbf{R}(\mathbf{p})\big)\cdot\mathbf{Q}(\mathbf{q})\cdot\mathbf{R}(\mathbf{q}),\\ H^{\mathbf{R}(\mathbf{p})}: &\quad \mathbf{q} \longmapsto f_{\mathbf{R}}\big(\mathbf{R}(\mathbf{q}),\mathbf{R}(\mathbf{p})\big)\cdot\mathbf{Q}(\mathbf{q}). \end{aligned} \tag{5}$$

We then may rewrite (3) as follows:

$$\begin{aligned} \mathbf{J_4}(\mathbf{p}) \;=\; & \big(1-\mathbf{Q}(\mathbf{p})\big)\frac{\sum_{\mathbf{q}\in N(\mathbf{p})}\big[f_{\mathbf{S}}(\mathbf{p},\mathbf{q})\cdot E^{\mathbf{I}(\mathbf{p})}(\mathbf{q})\big]}{\sum_{\mathbf{q}\in N(\mathbf{p})}\big[f_{\mathbf{S}}(\mathbf{p},\mathbf{q})\cdot F^{\mathbf{I}(\mathbf{p})}(\mathbf{q})\big]} + \\ & \mathbf{Q}(\mathbf{p})\frac{\sum_{\mathbf{q}\in N(\mathbf{p})}\big[f_{\mathbf{S}}(\mathbf{p},\mathbf{q})\cdot G^{\mathbf{R}(\mathbf{p})}(\mathbf{q})\big]}{\sum_{\mathbf{q}\in N(\mathbf{p})}\big[f_{\mathbf{S}}(\mathbf{p},\mathbf{q})\cdot H^{\mathbf{R}(\mathbf{p})}(\mathbf{q})\big]}. \end{aligned} \tag{6}$$

We note that $f_{\mathbf{S}}(\mathbf{p},\mathbf{q})$ is a function of the difference $(\mathbf{p}-\mathbf{q})$. We may hence write (6) as:

$$\mathbf{J_4}(\mathbf{p}) = \big(1-\mathbf{Q}(\mathbf{p})\big)\frac{\Big(f_{\mathbf{S}} * E^{\mathbf{I}(\mathbf{p})}\Big)(\mathbf{p})}{\Big(f_{\mathbf{S}} * F^{\mathbf{I}(\mathbf{p})}\Big)(\mathbf{p})} + \mathbf{Q}(\mathbf{p})\frac{\Big(f_{\mathbf{S}} * G^{\mathbf{R}(\mathbf{p})}\Big)(\mathbf{p})}{\Big(f_{\mathbf{S}} * H^{\mathbf{R}(\mathbf{p})}\Big)(\mathbf{p})}, \tag{7}$$

where $*$ denotes the convolution between functions. In addition and according to Paris et al. [12], the sampling of the data to be filtered does not introduce significant errors. In contrast, that ensures a good memory and speed performances. In [12], the authors present a study done on a set of images considering different $\sigma_{\mathbf{S}}$ and $\sigma_{\mathbf{I}}$ values as well as several sampling rates $(s_{\mathbf{S}}, s_{\mathbf{I}})$. They end with a consistent approximation when using a sampling rate proportional to the Gaussian bandwidth. The reformulation of the filter in (7) using two convolutions together with a sampling of the data to be filtered enable its implementation to perform in real-time, as we present in Section 5.3.

5. Experimental Results

In the following we analyse three main aspects of our depth enhancement filter. We first quantify our method against alternative filtering solutions. To that end, we consider our own recorded sequence as well as various scenes from the Middelbury dataset[1]. Then, we check the filter response against noise and we end with a runtime analysis using the filter implementation proposed in Section 4.

5.1. Quantitative comparison

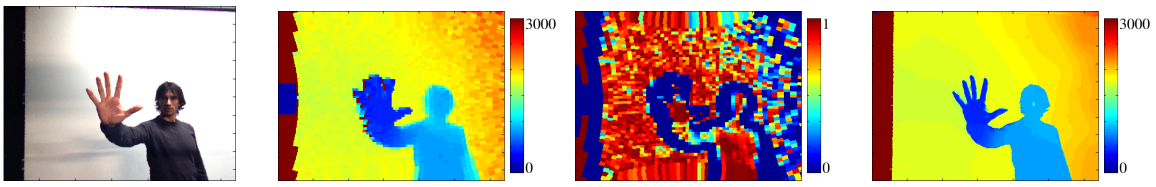
We start the assessment of our method with a qualitative comparison against the JBU and the PWAS filters employing real data. The test rig we have used comprises a 3D MLI Sensor™ from IEE S.A. [2] and a Flea®2 video camera from Point Grey™[3]. Both sensors are coupled with a narrow baseline of 30 mm. Also, they are calibrated for a perfect data alignment and frame-synchronised. Whereas the Flea®2 video camera provides (648×488) pixels, the 3D MLI Sensor™ provides a lower resolution of (56×61) pixels. Figure 1 shows the final depth map for two real scenarios in addition to the input data to be filtered, *i.e.,* the high-resolution 2-D image, the low-resolution depth map and the credibility map. First, one recognizes that our adaptive filter enhances the low-resolution depth maps from (56×61) pixels to the VGA-resolution of the coupled 2-D image. Also, the noise level has been greatly reduced. From the credibility map, depth edges weighted with a lower value, *i.e.,* 0, are accurately adjusted to the ones in the guidance image. Hence, resolving details like the fingers of the person in Figures 1(d), 1(h), that are not resolved in the raw depth map. Figure 2 compares a detailed region of our enhanced depth maps with the ones given by the JBU and the PWAS filters. In the first example one recognizes the edge blurring within the contour of the hand when filtering with JBU, which is drastically reduced for both the PWAS and the proposed filter. Although PWAS performance is not improved when adjusting depth edges, depth accuracy for pixels with a high credibility weight is increased by maintaining smooth regions. Figure 2 shows an example where the black belt of the person has the same (black) colour as the background. Contrary to the JBU and PWAS responses, our adaptive filter correctly addresses that situation, as shown in Figure 2(f).

In order to quantify our method against the alternative filtering solutions, we employ the *Teddy*, *Art*, *Books* and, *Moebius* scenes from the Middlebury dataset. Each scene contains an intensity image and its corresponding disparity map, from which we have generated a depth map as a ground truth using the also provided system specifications.

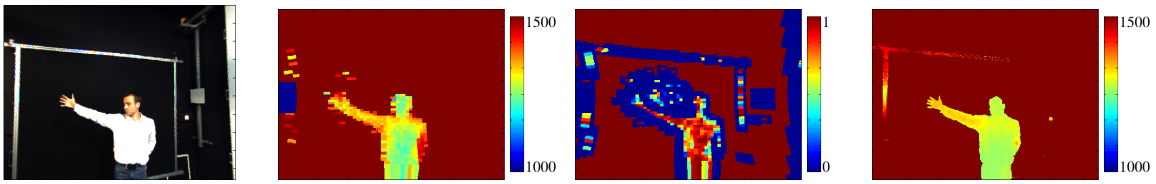
[1] Middlebury Stereo Dataset, http://vision.middlebury.edu/stereo/data

[2] IEE S.A., 3D MLI Sensor™, http://www.iee.lu

[3] Point Grey™, Flea®2, http://www.ptgrey.com/products/flea2/

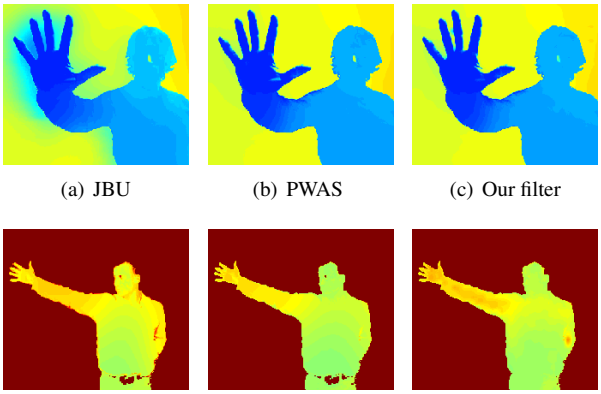


(a) 2-D guidance image (b) Low-resolution depth map (c) Credibility map (d) Enhanced depth map



(e) 2-D guidance image (f) Low-resolution depth map (g) Credibility map (h) Enhanced depth map

Figure 1. Depth map enhancement on our own recording sequences.



(a) JBU (b) PWAS (c) Our filter

(d) JBU (e) PWAS (f) Our filter

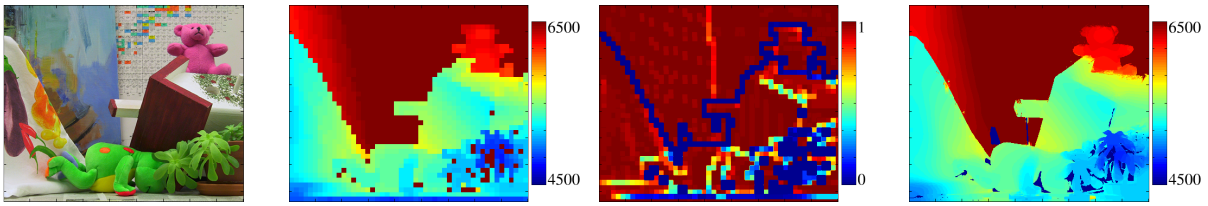
Figure 2. Visual comparison of enhanced depth maps using different depth enhancement filters.

We simulate the low-resolution depth map to be enhanced by downsampling (at different sampling rates) the ground truth depth map. Figure 3 shows an example of the *Teddy* and the *Art* scenes where the ground truth depth maps were downsampled by a factor of nine. As also occurs in the real data examples, our filter enhances the downsampled depth map to the intensity image resolution. Figure 4 shows a zoomed area where we can observe the same differences between the different filters applied on the real data examples shown in Figure 2. The JBU filter shows a strong edge blurring where the grey image contrast is low, *e.g.,* around the teddy's ears, marked as red rectangles in Figure 4(c). In addition to strongly reducing this artefact, our adaptive filter also removes the texture copying effect inside the teddy's head (see the green marked rectangle in Figure 4(c)), which remains in both JBU and PWAS final depth maps. Figure 5 shows an example of the limitations of our filter. This scene contains really small objects (in the regions indicated by red rectangles in 5(c)) that are tackled as outliers. This occurs because the credibility map gives a low weight to these objects and consequently their value is replaced by the neighbourhood pixel values. Exactly the same occurs when filtering using PWAS. However, on the larger surfaces in the scene (see areas inside the green rectangles in 5(c)), the resulting depth values of our filter are much more accurate than those of JBU and thus, on average, a better performance can be expected.

Although the root mean square error (RMSE) is a frequently-used measure to quantify the visibility of errors between a treated image and a reference image, we use an alternative complementary framework for quality assessment based on the degradation of structural information, the Structural SIMilarity (SSIM) Index [18]. Table 1 reports the SSIM measure that quantifies our method against alternative depth enhancement solutions. We can observe that under global error measure, our filter performs at least as well as the PWAS filter. The only case where our filter is not outperforming the JBU is in the *Art* scene with a downsampling rate of 3. This occurs due to the suppressed small details in the scene as discussed above. For higher downsampling rates, the performance is, however, superior to JBU.

5.2. Robustness to noise

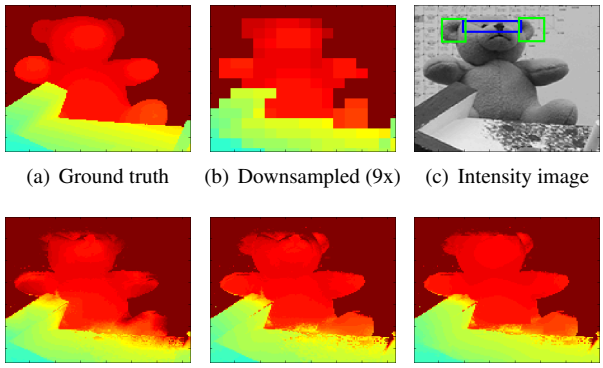
As shown in Figures 1(b), 1(f), ToF data is generally affected by random noise. We thus want to quantify how our filter behaves against different noise levels. Due to the active illumination of ToF cameras, the noise level increases according to the measured distance. Therefore we simulate this behaviour by adding Gaussian noise with a variance linearly dependent on the distance measurement [10]. We used the *Teddy* scene downsampled by a factor of 5 and with a noise of $\pm 100mm$ at the maximum distance (8976 mm). The results in the graph from Figure 6 were obtained by a Monte Carlo simulation over 100 times, which gave us an



(a) 2-D guidance image (b) Downsampled (9x) depth map (c) Credibility map (d) Enhanced depth map

(e) 2-D guidance image (f) Downsampled (9x) depth map (g) Credibility map (h) Enhanced depth map

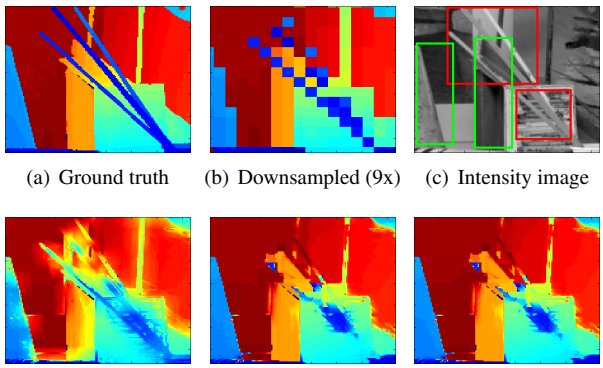
Figure 3. Depth map enhancement employing the *Teddy* and the *Art* scenes, 1^{st} and 2^{nd} rows respectively.



(a) Ground truth (b) Downsampled (9x) (c) Intensity image

(d) JBU output, SSIM: 62.62 (e) PWAS output, SSIM: 69.14 (f) Our output, SSIM: 69.95

Figure 4. Visual comparison employing the *Teddy* scene.



(a) Ground truth (b) Downsampled (9x) (c) Intensity image

(d) JBU output, SSIM: 44.01 (e) PWAS output, SSIM: 49.95 (f) Our output, SSIM: 50.13

Figure 5. Visual comparison employing the *Art* scene.

accuracy of $\pm 1.2 \times 10^{-3}$, $\pm 2.2 \times 10^{-4}$, and $\pm 2.2 \times 10^{-4}$ for the JBU, PWAS, and our filter, respectively. Within individual executions only the last digit varies.

Table 1. Quantitative comparison using the SSIM measure (100 corresponds to a perfect matching with the ground truth).

	Downsampled	JBU	PWAS	Our method
Teddy	3x	97.65	97.71	97.81
	5x	96.29	96.80	96.90
	9x	93.47	94.57	94.79
Moebius	3x	96.57	96.65	96.71
	5x	94.67	94.68	94.75
	9x	90.75	90.96	91.45
Books	3x	96.89	97.44	97.46
	5x	95.59	96.11	96.13
	9x	92.51	93.01	93.59
Art	3x	92.96	91.52	91.59
	5x	88.42	88.07	88.21
	9x	81.09	83.28	83.42

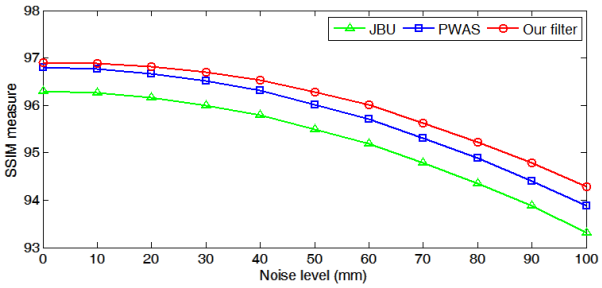


Figure 6. Filter responses against to Gaussian noise.

5.3. Runtime analysis

We next present a runtime analysis to validate that the implementation proposed in Section 4 enables for real-time applications. We ran the tests to estimate the time consumption on an Intel Core 2 Solo processor SU3500 (1.4 GHz, 800 MHz FSB) with an integrated graphic card Intel GMA 4500MHD. The filter was implemented in C language and

the tests have been performed on our own recorded scenes, enhancing from (56×61) pixels to VGA-resolution. Table 2 reports the seconds per filtered frame calculated over 1000 iterations. Also, we have sampled the input data by a factor of 2x, 4x, 8x and, 16x. With the latter sampling factor, the filtering process only takes 78 ms per frame. In addition, we have quantified the corresponding induced error to each sampling rate. Table 3 reports the SSIM measure considering the non downsampled case as a reference, and the final depth maps for each sampling rate. We notice that a sampling factor of 8x or 16x drastically reduces the time consumption without inducing a significant error in the final depth map. As a consequence, data sampling enables a real-time depth enhancement despite being restricted by the ToF camera frame rate of 11fps.

Table 2. Runtime analysis for the tested input data sampling rates (units are in seconds; average over 1000 iterations).

Sampling	JBU	PWAS	Our method
1x	1.88	1.89	13.59
2x	0.49	0.50	3.17
4x	0.13	0.13	0.65
8x	0.06	0.06	0.18
16x	0.05	0.05	0.08

Table 3. SSIM measure depending on the input data sampling.

Sampling	JBU	PWAS	Our method
2x	95.78	99.71	99.85
4x	95.46	99.51	99.65
8x	94.89	98.80	98.86
16x	92.25	95.11	95.17

6. Conclusion

In this paper, we have presented a new multi-lateral filter for low-level data fusion in real-time. The method enhances the low-resolution depth maps delivered by common ToF cameras up to the image resolution delivered by a coupled 2-D video camera in the hybrid ToF multi-camera rig. The generated dense depth maps present more accurate measurements where the depth discontinuities are well defined and adjusted to the 2-D guidance image. We increase the depth accuracy in such areas that are geometrically smooth, which are determined by the credibility map, adjusting the right weights within the filtering process. As a consequence of being based upon a bilateral filter, the filtered depth measurements are smoothed. Therefore, the global noise level is significantly reduced. The experimental results show that our filter outperforms previous fusion techniques, delivering better results even in the case where depth edges have no corresponding 2-D edges in the guidance image. In addition, we have proposed a fast implementation inspired from the work of Yang et al. [16] and following the recommendations of Paris et al. [12] that enables for real-time applications.

References

[1] D. Chan, H. Buisman, C. Theobalt, and S. Thrun. A noise-aware filter for real-time depth upsampling. In *ECCVW*, 2008.

[2] R. Crabb, C. Tracey, A. Puranik, and J. Davis. Real-time foreground segmentation via range and color imaging. In *CVPRW*, pages 1–5, 2008.

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

[4] S. Foix, G. Aleny, and C. Torras. Exploitation of time-of-flight (ToF) cameras. Technical Report IRI-TR-10-07, Institut de Robtica i Informtica Industrial, CSIC-UPC, 2010.

[5] A. Frick, B. Bartczack, and R. Koch. 3D-TV LDV content generation with a hybrid ToF-multicamera RIG. In *3DTV-CON*, pages 1–4, 2010.

[6] F. Garcia, B. Mirbach, B. Ottersten, F. Grandidier, and A. Cuesta. Pixel Weighted Average Strategy for Depth Sensor Data Fusion. In *ICIP*, pages 2805–2808, 2010.

[7] S. Gloud, P. Baumstarck, M. Quigley, Y. N. Andrew, and K. Daphne. Integrating visual and range data for robotic object detection. In *ECCVW*, 2008.

[8] A. Kolb, E. Barth, R. Koch, and R. Larsen. Time-of-Flight Sensors in Computer Graphics. In *Eurographics - State of the Art Reports*, pages 119–134, 2009.

[9] J. Kopf, M. Cohen, D. Lischinski, and M. Uyttendaele. Joint bilateral upsampling. In *SIGGRAPH*, page 96, 2007.

[10] R. Lange and P. Seitz. Solid-State Time-of-Flight Range Camera. *Journal of Quantum Electronics*, 37:390–397, 2001.

[11] K. Natroshvili, M. Schmid, M. Stephan, A. Stiegler, and T. Schamm. Real time pedestrian detection by fusing pmd and cmos cameras. In *Intelligent Vehicles Symposium*, pages 925–929, 2008.

[12] S. Paris and F. Durand. A fast approximation of the bilateral filter using a signal processing approach. In *International Journal of Computer Vision*, volume 81, pages 24–52, 2009.

[13] F. Porikli. Constant time o(1) bilateral filtering. In *CVPR*, pages 1–8, 2008.

[14] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *ICCV*, pages 839–846, 1998.

[15] Y. Wei, F. Franchetti, J. Hoe, C. Yao-Jen, and C. Tsuhan. Fast bilateral filtering by adapting block size. In *ICIP*, pages 3281–3284, 2010.

[16] Q. Yang, K.-H. Tan, and N. Ahuja. Real-time o(1) bilateral filtering. In *CVPR*, pages 557–564, 2009.

[17] Q. Yang, R. Yang, J. Davis, and D. Nistér. Spatial-depth super resolution for range images. In *CVPR*, pages 1–8, 2007.

[18] W. Zhou, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

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