

# Depth Enhancement by Fusion for Passive and Active Sensing<sup>\*</sup>

Frederic Garcia<sup>1</sup>, Djamila Aouada<sup>1</sup>, Hashim Kemal Abdella<sup>1,2</sup>,  
Thomas Solognac<sup>3</sup>, Bruno Mirbach<sup>3</sup>, and Björn Ottersten<sup>1</sup>

<sup>1</sup> Interdisciplinary Centre for Security, Reliability and Trust  
University of Luxembourg

{frederic.garcia,djamila.aouada,bjorn.ottersten}@uni.lu

<sup>2</sup> Université de Bourgogne

hashim-kemal\_abdella@etu.u-bourgogne.fr

<sup>3</sup> Advanced Engineering - IEE S.A.

{thomas.solognac,bruno.mirbach}@iee.lu

**Abstract.** This paper presents a general refinement procedure that enhances any given depth map obtained by passive or active sensing. Given a depth map, either estimated by triangulation methods or directly provided by the sensing system, and its corresponding 2-D image, we correct the depth values by separately treating regions with undesired effects such as empty holes, texture copying or edge blurring due to homogeneous regions, occlusions, and shadowing. In this work, we use recent depth enhancement filters intended for Time-of-Flight cameras, and adapt them to alternative depth sensing modalities, both active using an RGB-D camera and passive using a dense stereo camera. To that end, we propose specific masks to tackle areas in the scene that require a special treatment. Our experimental results show that such areas are satisfactorily handled by replacing erroneous depth measurements with accurate ones.

**Keywords:** depth enhancement, data fusion, passive sensing, active sensing.

## 1 Introduction

The demand to achieve an autonomous system that is capable of understanding the shape and location of objects in a scene has been growing in recent years. Hence the demand for a quality depth estimation is today one of the active research areas in computer vision. Triangulation methods are commonly used for depth perception, either using stereopsis or stereo vision [1]; the classic implementation of passive sensing, or in the case of active sensing by laser or structured light techniques [2]. It is known that regardless whether the sensing system is passive or active, triangulation methods can be quite time consuming

---

\* This work was supported by the National Research Fund, Luxembourg, under the CORE project C11/BM/1204105/FAVE/Ottersten.

as they have to cope either with the correspondence problem in the first case [1], or to process several encoded illumination patterns in the second case. Current advances in technology have greatly helped to significantly overcome this problem and stereo cameras such as the Bumblebee<sup>®</sup> XB3 from Point Grey Research or consumer depth cameras such as the recently appeared Microsoft Kinect are able to provide high-resolution depth maps in real-time. However, such sensing systems are based on triangulation techniques and thus, they are linked to the baseline between the two cameras or camera plus light source, which yields to occlusions or shadowing, and creates erroneous regions during depth estimation.

We propose to look into approaches of fusion by filtering tested and proven on Time-of-Flight (ToF) cameras [3–6]. Our main goal is to generalize such approach and define a global framework that may be adapted to other active sensors, specifically an RGB-D consumer camera, but also to more traditional passive ones, such as a dense stereo camera. The filtering is based on the concept of fusing the depth map with a guidance or a reference image (or images), usually taken as the matching 2-D image. This guidance image is used to correct unreliable depth regions. In this paper, we will design confidence measures, adequate to the considered sensor, to incorporate to the filter in order to indicate those areas within the initial depth map that require special attention.

The remainder of the paper is organized as follows: Section 2 presents the general framework of depth enhancement by fusion filters. In Section 3, we show how to apply these same filters to a stereo camera, then to a consumer RGB-D camera; hence, illustrating how this depth enhancement framework is generic. Section 4, presents the results of the proposed depth enhancement approach by data fusion in both modalities. Finally, concluding remarks are given in Section 5.

## 2 Problem Statement and Background

The idea of considering a guidance 2-D image to improve the quality of its corresponding depth map was first introduced by Kopf *et. al* in [7], where they presented the Joint Bilateral Upsampling (JBU) filter, an extension of the bilateral filter [8] that considers two different data sources within the kernel of the filter. Their work was first intended to compute a solution for image analysis and enhancement tasks, such as tone mapping or colourization through a downsampled version of the data. This idea was later applied for depth map enhancement in the context of real-time matting as presented by Crabb *et al.* [9]. The JBU filter enhances an initial depth map  $\mathbf{D}$  to the higher resolution of a corresponding 2-D guidance image  $\mathbf{I}$ , as follows

$$\mathbf{J}_1(\mathbf{p}) = \frac{\sum_{\mathbf{q} \in N(\mathbf{p})} f_s(\mathbf{p}, \mathbf{q}) f_I(\mathbf{I}(\mathbf{p}), \mathbf{I}(\mathbf{q})) \mathbf{D}(\mathbf{q})}{\sum_{\mathbf{q} \in N(\mathbf{p})} f_s(\mathbf{p}, \mathbf{q}) f_I(\mathbf{I}(\mathbf{p}), \mathbf{I}(\mathbf{q}))}, \quad (1)$$

where  $N(\mathbf{p})$  is the neighbourhood at the pixel indexed by the position vector  $\mathbf{p} = (i, j)^T$ , with  $i$  and  $j$  indicating the row, respectively column corresponding to the pixel position. This non-iterative filter formulation is a weighted average

of the local neighbourhood samples, where the weights are computed based on spatial and radiometric distances between the centre of the considered sample and the neighbouring samples. Thus, its kernel is decomposed into a spatial weighting term  $f_{\mathbf{S}}(\cdot)$  that applies to the pixel position  $\mathbf{p}$ , and a range weighting term  $f_{\mathbf{I}}(\cdot)$  that applies to the pixel value  $\mathbf{I}(\mathbf{q})$ . The weighting functions  $f_{\mathbf{S}}(\cdot)$  and  $f_{\mathbf{I}}(\cdot)$  are generally chosen to be Gaussian functions with standard deviations  $\sigma_{\mathbf{S}}$  and  $\sigma_{\mathbf{I}}$ , respectively. 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 2-D texture into actually smooth geometries within the depth map. Furthermore, a second unwanted artefact known as edge blurring appears along depth edges that have no corresponding edges in the 2-D image, *i.e.*, in situations where objects on either side of a depth discontinuity have a similar colour. In order to cope with these issues, Garcia *et. al* [4] proposed a new fusion filter known as Pixel Weighted Average Strategy (PWAS). The PWAS filter extends the expression in (1) by an additional factor, to which they refer as the credibility map, that indicates unreliable regions within the depth maps obtained using a Time-of-Flight (ToF) camera. These regions require a special treatment. Thus, for a given depth map  $\mathbf{D}$ , a credibility map  $\mathbf{Q}$ , and a guiding intensity image  $\mathbf{I}$ , the enhanced depth map  $\mathbf{J}_2$  resulting from PWAS filtering is defined as follows

$$\mathbf{J}_2(\mathbf{p}) = \frac{\sum_{\mathbf{q} \in N(\mathbf{p})} f_{\mathbf{S}}(\mathbf{p}, \mathbf{q}) f_{\mathbf{I}}(\mathbf{I}(\mathbf{p}), \mathbf{I}(\mathbf{q})) \mathbf{Q}(\mathbf{q}) \mathbf{D}(\mathbf{q})}{\sum_{\mathbf{q} \in N(\mathbf{p})} f_{\mathbf{S}}(\mathbf{p}, \mathbf{q}) f_{\mathbf{I}}(\mathbf{I}(\mathbf{p}), \mathbf{I}(\mathbf{q})) \mathbf{Q}(\mathbf{q})}. \quad (2)$$

with

$$\mathbf{Q}(\mathbf{p}) = \exp\left(\frac{-(\nabla \mathbf{D}(\mathbf{p}))^2}{2\sigma_{\mathbf{Q}}^2}\right), \quad (3)$$

where  $\nabla \mathbf{D}$  is the gradient of the given depth map  $\mathbf{D}$ . Although the PWAS filter copes well with the edge blurring artifact, texture copying is still not fully solved within the enhanced depth maps. In order to significantly reduce this artifact, the same authors proposed in [5] the Unified Multi-Lateral (UML) filter. The UML filter combines two PWAS filters where the output  $\mathbf{J}_3$  of the second one has both spatial and range kernels acting onto the same data source  $\mathbf{D}$ . In addition, they suggested to use the credibility map  $\mathbf{Q}$  as a blending function, *i.e.*,  $\beta = \mathbf{Q}$ , hence, depth pixels with high reliability are not influenced by the 2-D data avoiding texture copying as follows

$$\mathbf{J}_4(\mathbf{p}) = (1 - \beta(\mathbf{p})) \cdot \mathbf{J}_2(\mathbf{p}) + \beta(\mathbf{p}) \cdot \mathbf{J}_3(\mathbf{p}), \quad (4)$$

and

$$\mathbf{J}_3(\mathbf{p}) = \frac{\sum_{\mathbf{q} \in N(\mathbf{p})} f_{\mathbf{S}}(\mathbf{p}, \mathbf{q}) f_{\mathbf{D}}(\mathbf{D}(\mathbf{p}), \mathbf{D}(\mathbf{q})) \mathbf{Q}(\mathbf{q}) \mathbf{D}(\mathbf{q})}{\sum_{\mathbf{q} \in N(\mathbf{p})} f_{\mathbf{S}}(\mathbf{p}, \mathbf{q}) f_{\mathbf{D}}(\mathbf{D}(\mathbf{p}), \mathbf{D}(\mathbf{q})) \mathbf{Q}(\mathbf{q})}. \quad (5)$$

The kernel  $f_{\mathbf{D}}$  is another Gaussian function with a different standard deviation  $\sigma_{\mathbf{D}}$ .

### 3 Proposed Depth Enhancement Framework

It is well known that the estimation of depth using triangulation approaches entails to cope with the correspondence problem [1], *i.e.*, find feature-correspondence pairs between the two images in the case of passive systems, or detect the projected features in the case of active systems [2]. Problems in passive sensing arise when there are homogeneous regions in the scene, which prevents to find a correspondence between the two images or, in active sensing, when the light power of the projected pattern is not enough to be reflected back to the sensor. Furthermore, since both active and passive triangulation methods require a baseline for depth estimation, occlusion and shadowing are two additional drawbacks to overcome. In summary, depth maps obtained from triangulation approaches are unreliable in object boundaries, and in occluded or homogeneous regions. In order to deal with such regions, we propose to fuse these estimated depth maps with their corresponding 2-D images, *i.e.*, the images that were previously used to detect feature-correspondence pairs. In what follows, we propose to generalize state-of-the-art filters used in ToF depth enhancement and apply them to other passive and active sensors. This goes through identifying the regions that suffer from artifacts and require special treatment. Dealing with these regions requires defining the adequate credibility maps for each sensor.

#### 3.1 Application to Passive Sensing

We consider a stereo setup for passive sensing with two initial 2-D images; left image  $\mathbf{I}_l$  and right image  $\mathbf{I}_r$ . The depth map  $\mathbf{D}$  is defined by finding the dense correspondence between  $\mathbf{I}_l$  and  $\mathbf{I}_r$ . We adapt the depth enhancement filters presented in Section 2 to depth maps obtained using passive sensing by considering different kinds of credibility maps or masks. Similarly to ToF depth maps, object boundaries are represented by a *boundary map*  $\mathbf{Q}_b$ , which corresponds to the credibility map in (3) ( $\mathbf{Q}_b = \mathbf{Q}$ ). However, the edges in the estimated depth map  $\mathbf{D}$  may be misaligned with their corresponding edges in the 2-D image. Therefore, we propose to dilate  $\mathbf{Q}_b$  according to the gradient strength. To that end, we first bin the gradient into five levels. Then, pixels with the highest gradient are dilated strongly while those with the lowest gradient undergo a minimum dilation. Bin centres and levels of dilation are manually tuned.

The second type of credibility map is the *occlusion map*  $\mathbf{Q}_o$ , which is determined through left/right consistency check [10]. In left/right consistency check, after obtaining the disparity relative to the right image, a comparison is made between the left disparity  $\mathbf{D}_l$  and the right one  $\mathbf{D}_r$ . In this comparison, corresponding pixels are supposed to have the same value, and pixels which deviate from this assumption are considered as occluded in either views. Therefore, the occlusion mask  $\mathbf{Q}_o$  is defined as

$$\mathbf{Q}_o(\mathbf{p}) = \exp \left( \frac{-\|\mathbf{D}_l(\mathbf{p}) - \mathbf{D}_r(\mathbf{p} + \mathbf{D}_l(\mathbf{p}))\|^2}{2\sigma_o^2} \right). \quad (6)$$

Homogeneous regions are also another source of unreliability. In [11], the characteristics of correlation cost at each pixel is analysed for determining a *homogeneity mask*  $\mathbf{Q}_h$ . Accordingly, pixels in homogeneous regions have a flat correlation cost, while repetitive patterns give rise to a cost function with multiple minima. The cost function  $C$  is computed at  $\mathbf{p}$  for all possible disparity values. It has a first minimum value at depth  $d_1$  and a second minimum at  $d_2$ , with corresponding costs  $C(\mathbf{p}, d_1)$  and  $C(\mathbf{p}, d_2)$ , respectively. These costs are used to define  $\mathbf{Q}_h$  as

$$\mathbf{Q}_h(\mathbf{p}) = 1 - \exp\left(\frac{-\left(\frac{C(\mathbf{p}, d_2) - C(\mathbf{p}, d_1)}{C(\mathbf{p}, d_1)}\right)^2}{2\sigma_h^2}\right), \quad (7)$$

where  $\sigma_b$ ,  $\sigma_o$ , and  $\sigma_h$  are empirically defined parameters.

As a final credibility map for stereo depth enhancement, we propose the following combined one:

$$\mathbf{Q}(\mathbf{p}) = \mathbf{Q}_b(\mathbf{p}) \cdot \mathbf{Q}_o(\mathbf{p}) \cdot \mathbf{Q}_h(\mathbf{p}). \quad (8)$$

### 3.2 Application to Active Sensing

We herein consider a consumer RGB-D camera as an active sensor, where a depth map  $\mathbf{D}$  is acquired simultaneously to a perfectly matching 2-D image  $\mathbf{I}$ . In contrast to ToF cameras, depth measurements given by RGB-D cameras are known to be much less influenced by noise. This, in turn, allows us to avoid the second PWAS filtering  $\mathbf{J}_3$  in (4) as it can be directly replaced by the acquired depth map  $\mathbf{D}$ , *i.e.*,

$$\mathbf{J}_5(\mathbf{p}) = (1 - \beta(\mathbf{p})) \cdot \mathbf{J}_2(\mathbf{p}) + \beta(\mathbf{p}) \cdot \mathbf{D}(\mathbf{p}). \quad (9)$$

By doing so, reliable depth measurements are not smoothed and the complexity of the UML filter is comparable to the one of the PWAS filter, which guarantees a real-time performance. However, we realise that if we follow the recommendations of Garcia *et. al* [5] and we set the blending function  $\beta$  equal to the credibility map  $\mathbf{Q}$ , edge blurring will appear when filtering low reliable depth pixels if no 2-D edge is present. Indeed, this situation occurs when foreground and background objects share the same intensity value, which often occurs when considering grayscale images. Hence, we propose to not rely on the 2-D guidance image when depth measurements have a low reliability, *i.e.*,  $\mathbf{Q}_D < \tau_D$  ( $\mathbf{Q}_D = \mathbf{Q}$ , defined in (3)), and no corresponding 2-D edge, *i.e.*,  $\mathbf{Q}_I > \tau_I$ .  $\mathbf{Q}_I$  is defined analogously to  $\mathbf{Q}_D$  but considering  $\nabla \mathbf{I}$ . The constants  $\tau_I$  and  $\tau_D$  are empirically chosen thresholds. We therefore generalise the blending function  $\beta$  in (9) as follows

$$\beta(\mathbf{p}) = (u_I(\mathbf{p}) \cdot u_D(\mathbf{p})) + \mathbf{Q}_D(\mathbf{p}) \cdot (1 - u_I(\mathbf{p}) \cdot u_D(\mathbf{p})), \quad (10)$$

with

$$u_I(\mathbf{p}) = u(\mathbf{Q}_I(\mathbf{p}) - \tau_I) \quad \text{and} \quad u_D(\mathbf{p}) = u(\tau_D - \mathbf{Q}_D(\mathbf{p})), \quad (11)$$

being the function  $u(\cdot)$  a step function.

Though bilateral filtering is known to be time consuming, its latest fast implementations based on data quantization and downsampling [12, 13], also applicable to both PWAS and UML filters as demonstrated in [5], enable a high-performance. Thus, in order to ensure a real-time performance, we propose to downsample by a factor of  $s$  the credibility maps  $\mathbf{Q}_D$  and  $\mathbf{Q}_I$  preserving such regions that require special treatment. To that end, we keep the most significant pixel value within each downsampled block of size  $(s \times s)$  in the resulting low resolution image  $\mathbf{Q} \downarrow_s$ , *i.e.*,

$$\mathbf{Q} \downarrow_s(\mathbf{p}) = \min_{\mathbf{q}} \mathbf{Q}(\mathbf{q}) \text{ s.t. } \|\mathbf{q} - (\mathbf{p} + \mathbf{s})\| < \sqrt{2}s \text{ and } \mathbf{s} = (s, s)^T. \quad (12)$$

## 4 Experimental results

The proposed enhancement method has been evaluated using three main evaluation metrics: Root Mean Square (RMS), Percentage of Bad Matching Pixels (PBMP), and Structural SIMilarity (SSIM) [14, 15]. RMS is the simplest and the most widely used evaluation technique; yet, criticized for not representing the perceived visual quality [15]. We also found out that RMS value has an exaggeration when there are some pixels with a higher depth deviation. PBMP based evaluation is the common way of comparing depth estimation techniques. It envelopes a threshold to determine the quality of pixels' depth. Mostly this threshold is set to one, which makes any pixel with a deviation slightly greater than one to be set as bad. Our proposed fusion filter has a smoothing effect which will unarguably cause this deviation, making PBMP an inappropriate quality metric. Hence, we prefer to base our evaluation mostly on SSIM, which tries to compare two images based on the luminance, contrast and structural similarity [15].

### 4.1 Passive Sensing

The results in Table 1 present the quantitative evaluation of the proposed approach applied on the Teddy scene from the Middlebury dataset [16] shown in

**Table 1.** Quantitative evaluation using the three evaluation metrics on Teddy image sets

Depth map		Evaluation (best - bad)		
		RMS(0 - 1)	PBMP(0 - 100)	SSIM(1 - 0)
Initial		0.3732	13.9253	0.9689
PWAS	$\mathbf{Q}_h$	0.3974	15.7890	0.9842
	$\mathbf{Q}_b$	0.3792	14.3799	0.9843
	$\mathbf{Q}_o$	0.4468	19.9613	0.9844
	$\mathbf{Q}$	0.3688	13.5980	0.9849
JBU		0.4606	21.2127	0.9822

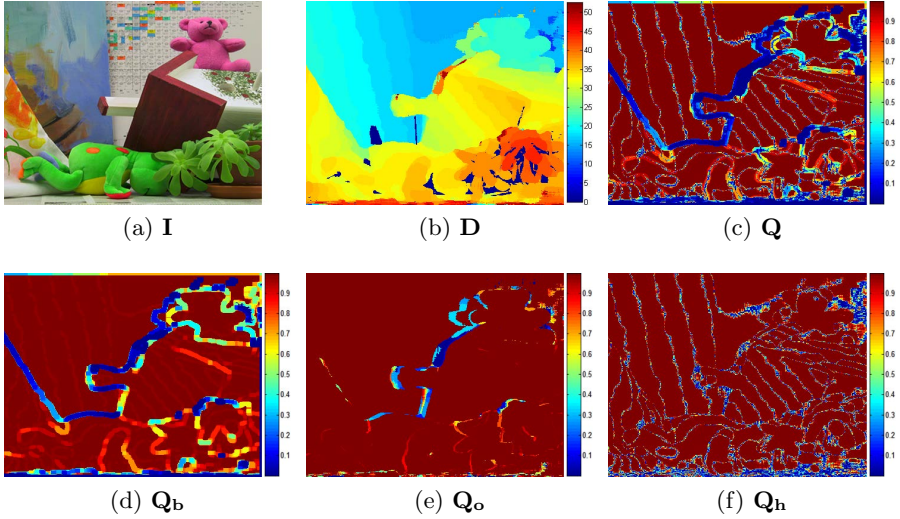


Fig. 1. Input data and proposed unreliability maps

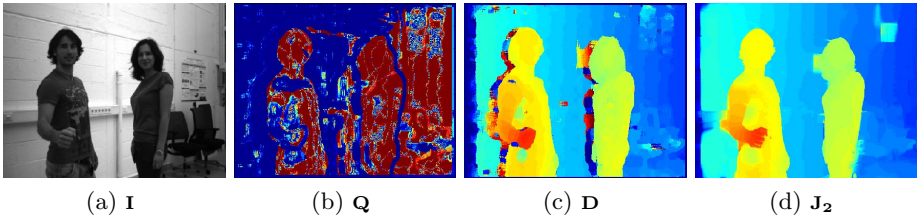


Fig. 2. Application of our proposed enhancement technique on stereo data acquired by a Bumblebee<sup>®</sup> XB3 camera

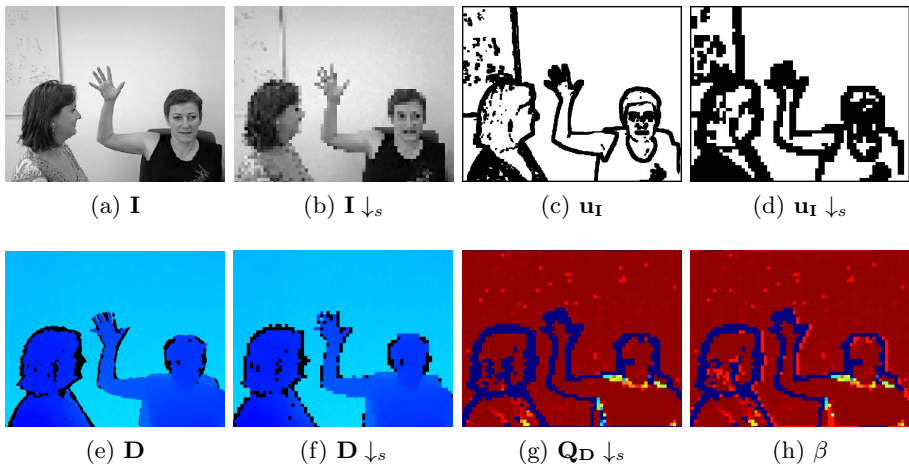
Fig. 1. The scene contains an intensity image and its corresponding disparity map, from which we have generated a depth map as a ground truth using the also provided system specifications. It is clear that PWAS filter using  $Q$  enhances the initial disparity (obtained using Matlab’s disparity function) in all the metrics. However, individual masks show improvements in one measure while there is a degradation in the other. We can also observe that the JBU output is lower than that of PWAS filter. Most importantly, the PBMP and RMS measures are very bad for JBU filter which is due to the smoothing effect of the filter in regions which were perfect in the initial disparity. This effect is slightly controlled by the credibility map in PWAS filtering.

We have also tested the proposed technique on our own recordings using a Bumblebee<sup>®</sup> XB3 camera [17]. Fig. 2 shows the result using one of these recordings. The reference image  $I$  is shown in Fig. 2a while the acquired depth map  $D$  is shown in Fig. 2c. The two people in  $D$  can be well identified without

much extension of boundaries; however, most of the occluded background regions (to the left of each person) and homogeneous regions (top right part of the image) appear to be unreliable. These regions are identified using the three masks defined in Section 3.1. The credibility map  $\mathbf{Q}$  shown in in Fig. 2b is the combination of these masks. Fig. 2d shows the enhanced depth map  $\mathbf{J}_2$  that results from the proposed fusion filter. As can be observed, the PWAS filter improves the given depth map except around the left person’s elbow, the top right of the image and next to the face of the person in the right. This is due to large holes created on the credibility map due to illumination variation and extended textureless surfaces. The PWAS filter can fill small holes, but it is impossible to handle a hole bigger than the fusion window.

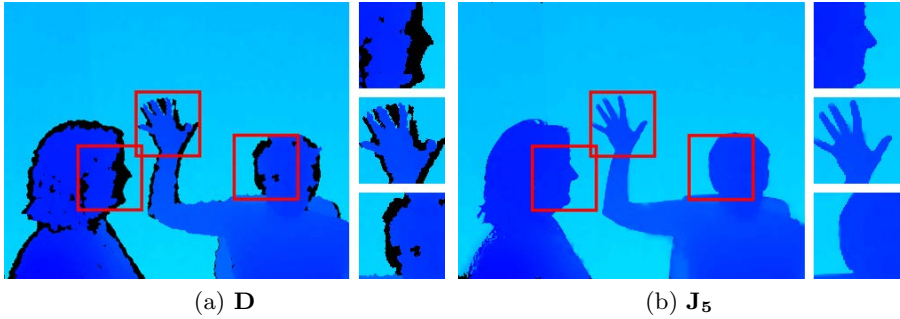
## 4.2 Active Sensing

For the evaluation of our approach on the modality of active sensing we have considered data captured using the Kinect camera. Fig. 3 is a visual example of the credibility maps proposed in Section 3.2 computed on the acquired depth map  $\mathbf{D}$  and the corresponding guidance image  $\mathbf{I}$ , shown in Fig. 3e and Fig. 3a, respectively. Fig. 3d is the downsampled version of  $\mathbf{u}_I$ , which indicates the presence of a 2-D edge while Fig. 3g corresponds to the downsampled version of  $\mathbf{Q}_D$ , indicating the presence of a depth edge. The proposed blending function  $\beta$  is shown in Fig. 3h. Note that low-reliability depth pixels in  $\mathbf{Q}_D \downarrow_s$  that have no correspondence 2-D edge in  $\mathbf{u}_I \downarrow_s$  has been setted to the maximum confidence value in  $\beta$ . The enhanced depth map  $\mathbf{J}_5$  that results from the proposed approach is shown Fig. 4b. Table 2 compares the SSIM evaluation of the proposed approach with state-of-the-art depth enhancement methods. For this comparison, we consider only those pixels from  $\mathbf{D}$  that have a defined depth value. Thus,



**Fig. 3.** Input data and proposed credibility and blending masks





**Fig. 4.** (a) Initial depth map acquired by the Kinect camera. Black areas indicate non-valid (occluded or shadow) pixels. (b) Enhanced depth map using the proposed approach in (9)

**Table 2.** Quantitative evaluation using the SSIM evaluation metric on the Kinect data presented in Fig. 3

	JBU filter	PWAS filter	UML filter	Proposed approach
SSIM(1 - 0)	0.73	0.83	0.83	0.86

occlusion or shadowing regions have not been considered. As can be observed, the depth measurements of the enhanced depth map using our approach are closer to the ones of the acquired depth map.

## 5 Conclusions

In this paper, we have showed how a depth enhancement technique designed for ToF depth maps may be generalized to other depth sensing modalities, both passive via a stereo camera and active via an RGB-D camera. Our main contribution was in identifying each time the areas of erroneous measurements. We defined accordingly credibility maps to enforce corrections on these areas only. However, by defining more credibility maps, more empirical parameters are introduced each time. This makes the practical automated usage of such fusion filter challenging. Another question that still needs investigation is a clear rule on when a smoothing by filtering is to be avoided and when a simple binary decision is to be undertaken as in (10).

**Acknowledgements.** We would like to thank Isabelle Chesnay and Cecile Petit, for posing in our experiments for active sensing.

## References

1. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd edn. Cambridge University Press (2003)

2. Salvi, J., Fernandez, S., Pribanic, T., Llado, X.: A state of the art in structured light patterns for surface profilometry. *Pattern Recognition* 43, 2666–2680 (2010)
3. 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, ECCVW* (2008)
4. Garcia, F., Mirbach, B., Ottersten, B., Grandidier, F., Cuesta, A.: Pixel Weighted Average Strategy for Depth Sensor Data Fusion. In: *International Conference on Image Processing (ICIP)*, pp. 2805–2808 (2010)
5. Garcia, F., Aouada, D., Mirbach, B., Solignac, T., Ottersten, B.: A New Multi-lateral Filter for Real-Time Depth Enhancement. In: *Advanced Video and Signal-Based Surveillance, AVSS* (2011)
6. Min, D., Lu, J., Minh, N.D.: Depth Video Enhancement Based on Weighted Mode Filtering. *IEEE Transactions on Image Processing (TIP)* 21, 1176–1190 (2012)
7. Kopf, J., Cohen, M., Lischinski, D., Uyttendaele, M.: Joint bilateral upsampling. In: *SIGGRAPH 2007: ACM SIGGRAPH 2007 papers*, p. 96. ACM, New York (2007)
8. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: *ICCV*, pp. 839–846 (1998)
9. Crabb, R., Tracey, C., Puranik, A., Davis, J.: Real-time foreground segmentation via range and color imaging. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–5 (2008)
10. Benhimane, S., Malis, E.: Real-time image-based tracking of planes using efficient second-order minimization. In: *IEEE IROS*, vol. 1, pp. 943–948 (2004)
11. Hirschmuller, H.: Real-time correlation-based stereo vision with reduced border errors. *IJCV* 47(1/2/3), 229–246 (2002)
12. Yang, Q., Tan, K.H., Ahuja, N.: Real-time  $O(1)$  bilateral filtering. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 557–564 (2009)
13. Paris, S., Durand, F.: A fast approximation of the bilateral filter using a signal processing approach. *International Journal of Computer Vision* 81, 24–52 (2009)
14. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision* 47, 7–42 (2002)
15. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: From error visibility to structural similarity. In: *IEEE TIP*, vol. 13-4, pp. 600–612 (2004)
16. Middlebury Stereo Datasets (2011), <http://vision.middlebury.edu/stereo/data/>
17. Point Grey Research, Inc. (2011), <http://www.ptgrey.com>