Supplementary Material
Depth-aware Motion Magnification

Julian F. P. Kooij[1,2] and Jan C. van Gemert[1,2]

[1] Delft University of Technology, Delft, The Netherlands
{J.F.P.Kooij,J.C.vanGemert}@tudelft.nl
[2] Leiden University Medical Center, Leiden, The Netherlands

This document provides additional experiments for our paper [1].

- Section 1 investigates the effect of depth image quality and alignment on motion magnification.
- Section 2 demonstrates other uses of the non-Gaussian bilateral filter, unrelated to motion magnification.

1 Effect of depth image quality and alignment

The paper describes that the depth images are *pre-processed* to remove artifacts. The pre-processing step consists of a applying 2D median and 2D max filter on the depth images. In this section we investigate how pre-processing affects the alignment with the color image, and how it removes noise. We first look at the effect of pre-processing on a single frame under various conditions, and then at how it affects the motion magnification results.

1.1 Single frame depth and color image alignment

Figures 1 and 2 compare the effect of pre-processing on a single frame from sequence 2 in our paper under various conditions. Each depth/color image pair demonstrates how the depth aligns with the color image by highlighting in red in the color image the 'foreground' region at a depth distance below 2.5 meters. Pairs in the top row use the original depth input, the bottom row the depth after pre-processing. The following conditions are shown:

Normal depth image First, 1a shows the original depth image. It contains some depth outliers (e.g. bright yellow blocks near armpit) and missing depth values (e.g. around the arm). These are natural effect of the depth sensor, but artifacts can also occur due to the depth-to-image projection, and upscaling of the low-res depth image to the color HD format. Without pre-processing (see top row), the foreground region is affected by noise and outliers and fits too tight around the arm, hence various border pixels of the arm fall in the background. The proposed pre-processing step (bottom row) removes the noise and fills most of the holes. The foreground region is also slightly enlarged such that the moving left arm is now fully contained.

Additional depth noise Then, 1b and 1c demonstrate the effect of adding zero-mean normally distributed noise, $\epsilon \sim \mathcal{N}(0, \sigma_n)$, to the original depth image, using low $\sigma_n = 50mm$, and high $\sigma_n = 100mm$ variance respectively. However, the bottom rows show that the proposed pre-processing yields almost identical results as the original input with respect to the moving arm.

Additional depth outliers Figures 2a and 2b consider additional extreme values. Since normally the depth-to-image projection upscales the depth image, an individual outlier can create a large block in the mapped depth image. We therefore add artificial outliers as 4×4 square regions with extreme values. In 2a, about $\sim 1\%$ of depth pixels is made an outlier, and in 2b this number is increased to $\sim 20\%$.

We can see that in the first case, pre-processing still removes the moderate amount of ouliers, thus results remain mostly unaffected. But, the second case shows that if there are too many outliers close together, the proposed pre-processing can result in fewer but even larger regions in the filtered result.

Temporal misalignment Finally, we consider the case that the depth and color frames are misaligned. Each depth frame t is therefore now combined with color frame $t + 3$, which increases the risk of the foreground inadvertently moving into the magnified region. At $\sim 30Hz$, 3 frames corresponds to about $1/10$ seconds. As can be seen in the top part of Figure 2c, the finger tips (which tremble with largest amplitude) are not contained in the foreground region anymore without pre-processing. The bottom part of the figure shows however that by expanding the foreground in the pre-processing step the hand is still fully contained, in this case.

We conclude that median filtering suffices to remove the noise in most conditions, and foreground expansion using the 2D max filter increases the robustness to spatial misalignment resulting from depth-to-color mapping, upscaling, and temporal offsets. Of course, if the motion of the foreground is too large, then the spatial misalignment resulting from a frame offset could be too big for the simple foreground expansion.

1.2 Effect on depth-aware motion magnification

Next, we magnify the same depth region of sequence 2 as described in the paper. Figure 3 illustrates again the temporal effects on a single vertical slice of pixels containing the trembling left arm (foreground which should remain unmagnifed), and the background which should be magnified. For reference, Figures 3a shows again the selected spatial slice, 3b the magnification matte, and 3c is the space-time slice of the original unmagnified video.

Figures 3d-3i demonstrate the effect of depth-aware motion magnification on the slice under the conditions explored in the previous section. For each

condition, the left space-time slice show the result without pre-processing, and the right slice with pre-processing.

As we found in the previous section, the pre-processing step ensures that the artifacts are mitigated in most conditions. In our test cases, the only exception is the condition containing 20% additional depth outliers, Figure 3h. As was shown in Figure 2b, these outliers are not fully removed by the used filtering strategy and the presence of these artifacts results in irregular and unrealistic magnification patterns, see 3h (right).

2 Other uses of non-Gaussian bilateral filtering

In the paper, the non-Gaussian bilateral filtering approach has only been demonstrated in the context of motion magnification. This section presents examples of some other possible image filtering applications.

2.1 Suppressing object edges from depth

First, we study extracting texture edges in a single RGB+D image taken from a frame of the Sintel dataset [2]. The input contains various occluding objects and an accurate depth map such that no depth pre-processing is needed, see the image in Figure 4a and its depth channel in Figure 4b. We consider using the depth channel to guide a bilateral filter on the color image.

While standard bilateral filtering with a Gaussian filter *removes texture edges* but *keeps edges at object boundaries* (see Figure 4c), our non-Gaussian bilateral filter can instead be used to *detect texture edges* while *suppressing edges at the object boundaries*. To demonstrate this, we filter the intensity of the RGB image with a steerable filter, and visualize in Figures 4d-4f the edge orientation (color coded, see color wheel in upper-left corner) and magnitude (saturation).

The effect of the depth on the filtered results can be adjusted by altering the depth range σ_r. In the extreme case with large σ_r, the non-Gaussian bilateral filter reduces to a standard non-Gaussian filter which does not use depth information, see Figure 4d. But, as σ_r decreases, the filter increasingly ignores intensity differences that coincide with depth differences, see Figure 4e ($\sigma_r = 1m \rightarrow 10$ depth layers) and 4f ($\sigma_r = 0.1m \rightarrow 100$ depth layers). Note how object edges in the black encircled regions are present in the standard filter, but removed by our bilateral non-Gaussian filter. On the other hand, texture edges (such as those on the bamboo and face in the white encircled regions) are kept since the depth around these edges is nearly uniform.

2.2 Non-Gaussian bilateral filtering with intensity only

Instead of relying on a related depth map to filter an input image, one could also consider using the intensity channel of the input itself as illustrated in Figure 5.

Figure 5a shows a grayscale input image. While the original bilateral filter performs edge-aware smoothing (Figure 5b), the non-Gaussian bilateral filter can

be used as an edge detector that ignores strong edges, emphasizing microtextures instead. On the input image, we build a normal steerable pyramid, and our proposed bilateral steerable pyramid using the intensity image itself to build the depth layers.

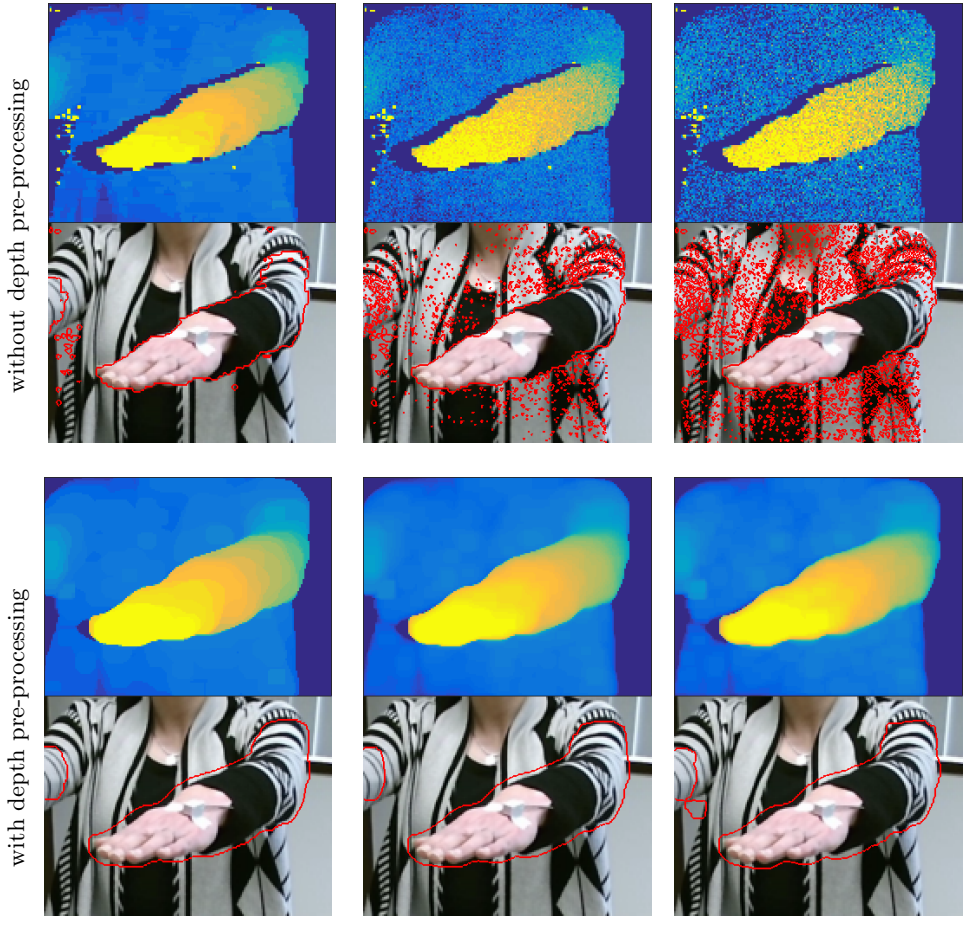
As in the previous subsection, we visualize the extract edge orientation, but now do not encode magnitude with saturation anymore as only low magnitude edges are kept.

In Figure 5c with the normal steerable pyramid, we see that large intensity differences dominate the edge orientation in a large spatial extent (determined by the size of the filter). For example, where the dark body outline contrasts against the lighter flat background, the computed orientation does not show any of the subtle details of the textures (e.g. inside the body, legs). The affected local neighborhood grows in size at subsequent pyramid levels, see Figure 5d.

In contrast, our bilateral non-Gaussian filter suppresses the influence of the large intensity differences, uncovering the texture orientation within regions even close to the object border, see Figures 5e and 5f. In Figures 5e for instance, one can see that more texture orientations in the legs and elbows than in 5c. Likewise, one can see more details in the body in 5f compared to 5d.

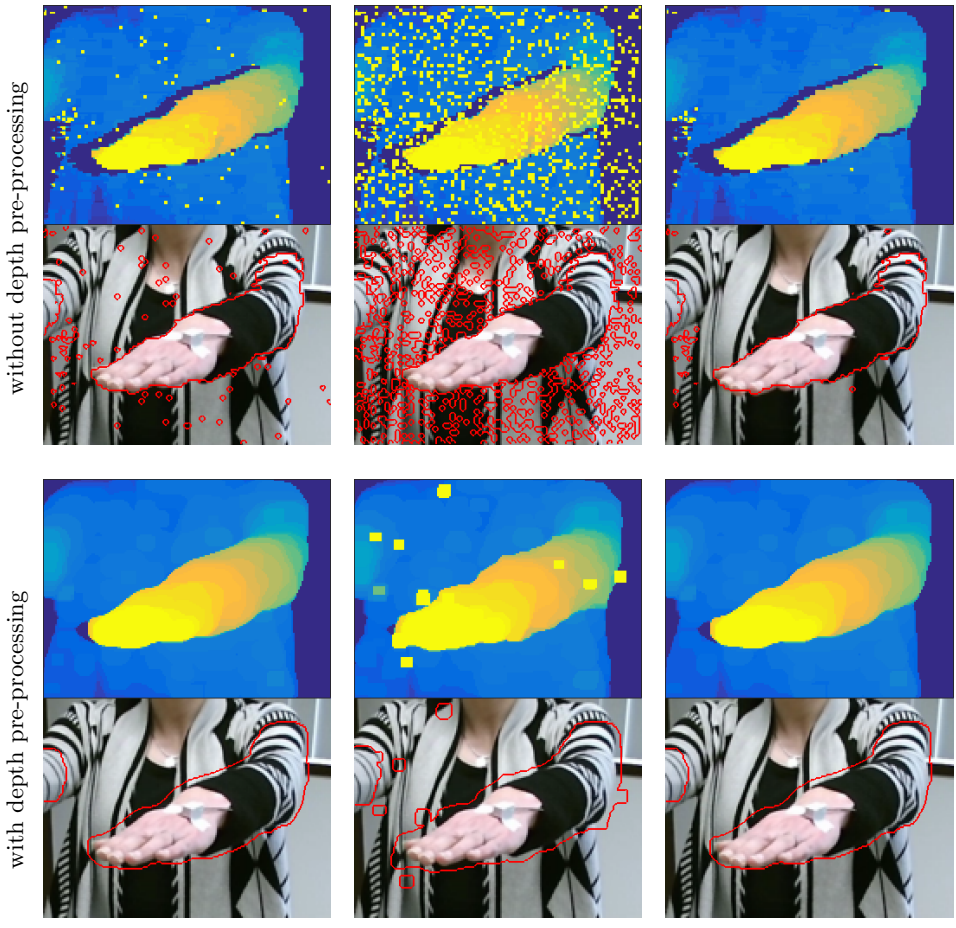
References

1. Kooij, J.F.P., van Gemert, J.C.: Depth-aware motion magnification. In: European Conference on Computer Vision (ECCV), 2016, Springer (to appear)
2. Butler, D.J., Wulff, J., Stanley, G.B., Black, M.J.: A naturalistic open source movie for optical flow evaluation. In A. Fitzgibbon et al. (Eds.), ed.: European Conf. on Computer Vision (ECCV). Part IV, LNCS 7577, Springer-Verlag (October 2012) 611–625



(a) proposed (b) white noise $\sigma_n = 50mm$ (c) white noise $\sigma_n = 100mm$

Fig. 1: Depth/color image pairs to demonstrate alignment, where red regions in color image corresponds to depth distance < 2.5 meter. Each column corresponds to different depth image quality, where top row is without depth pre-processing, and bottom row with. Comparing the bottom row, pre-processing ensures that the visual fore- and background are separably by depth in all cases.
(a) Original depth image as used in the paper. (b) With additional zero-mean Gaussian noise added to the depth $\sigma_n = 50mm$. (c) With more zero-mean Gaussian noise added to the depth $\sigma_n = 100mm$.

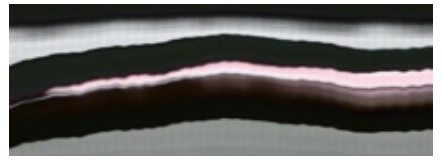


(a) 1% outliers (b) 20% outliers (c) temporal offset 3 frames

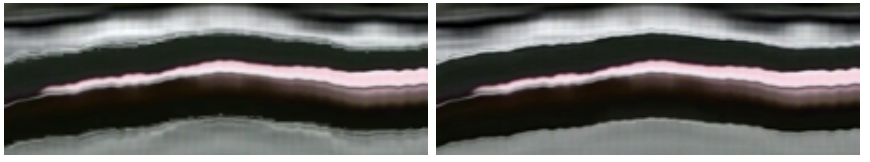
Fig. 2: Depth/color image pairs to demonstrate alignment, where red regions in color image corresponds to depth distance < 2.5 meter. Each column corresponds to different depth image quality, where top row is without depth pre-processing, and bottom row with. Comparing the bottom row, pre-processing ensures that the visual fore- and background are separably by depth, except in extreme cases with too many outliers.
(a) With about 1% of depth pixels set to extreme values, pre-processing still removes all artifacts. (b) With about 20% outliers pre-processing cannot remove all artifacts. (c) With temporal misalignment of 3 frames between depth and color image. Without pre-processing the fingertips of the trembling hand cross the foreground border.

(a) sample region

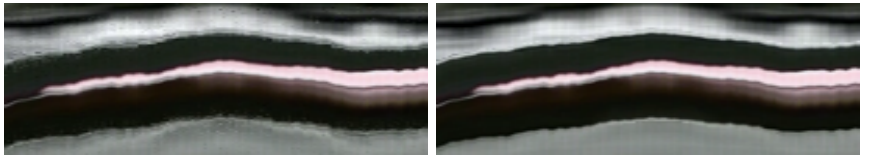
(b) magnify matte



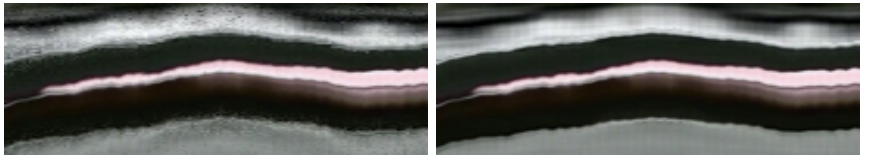
(c) no magnification



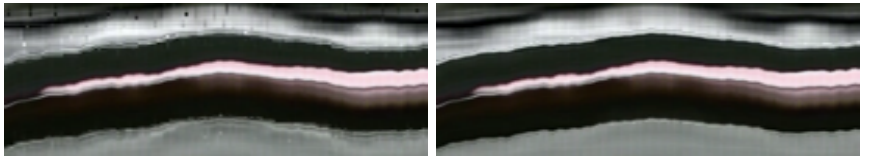
(d) proposed method (left: without, right: with pre-processing)



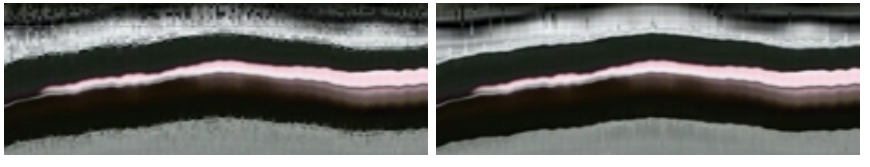
(e) depth white noise $\sigma = 50mm$ (left: without, right: with pre-processing)



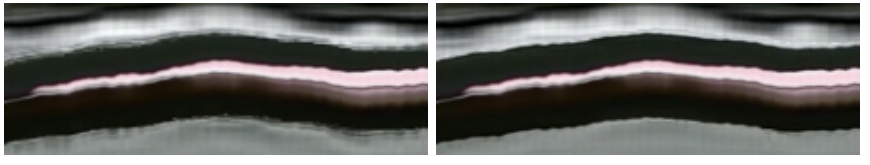
(f) depth white noise $\sigma = 100mm$ (left: without, right: with pre-processing)



(g) 1% depth outliers (left: without, right: with pre-processing)



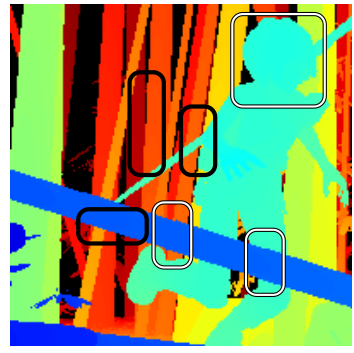
(h) 20% depth outliers (left: without, right: with pre-processing)



(i) temporal offset 3 frames (left: without, right: with pre-processing)

Fig. 3: Pre-processing the depth input ensures that motion magnification remains large unaffected, even with additional artificial depth errors and temporal offsets.

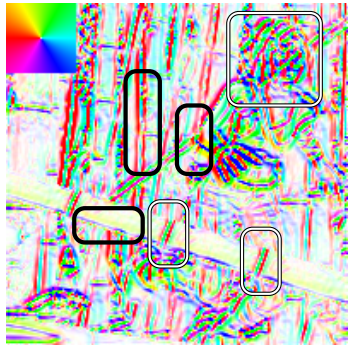
(a) input image "bamboo_2"



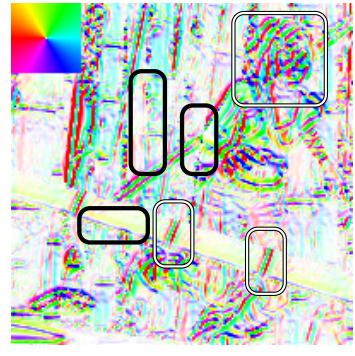
(b) depth image

(c) standard bilateral filter

(d) standard non-Gaussian filter



(e) bilateral non-Gaussian filter, $\sigma_r = 1$ meter

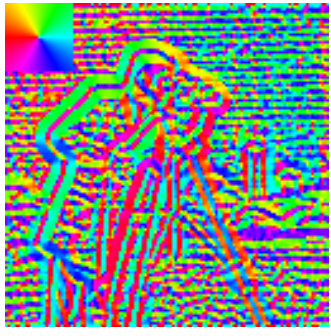


(f) bilateral non-Gaussian filter, $\sigma_r = 0.1$ meter

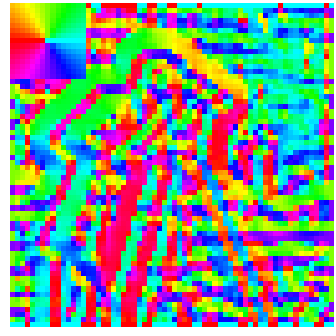
Fig. 4: (a) input frame from Sintel dataset [2], and (b) its depth frame. (c) A standard bilateral filter for edge-aware Gaussian blurring. (d) A standard steerable filter. Edges are color coded: orientation by color hue (see color-key in top-left corner), edge magnitude by saturation. (e), (f) Our method suppresses strong image edges that coincide with large depth differences (see *black* encircled regions), while keeping edges within objects where depth differences are small (see *white* encircled regions). The sensitivity for depth increases as σ_r decreases.

(a) input image, “cameraman”

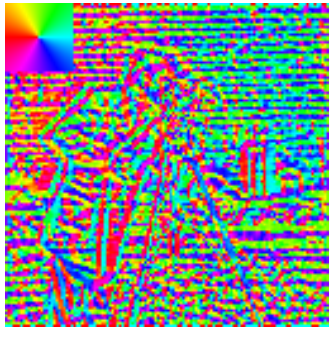
(b) standard bilateral Gauss blur



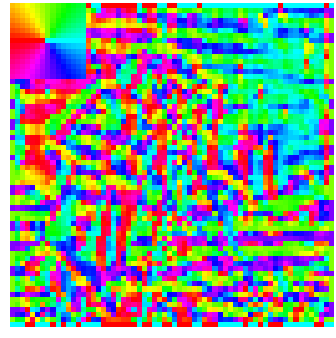
(c) standard pyramid, level 2



(d) standard pyramid, level 3



(e) bilateral pyramid, level 2 (ours)



(f) bilateral pyramid, level 3 (ours)

Fig. 5: Using intensity instead of depth for non-Gaussian bilateral filtering. (a) input image. (b) applying a standard bilateral filter blurs the image while maintaining strong edges. (c), (d) edge orientations at various levels (scales) in a standard steerable pyramid (upper-left corner displays the orientation color-key). (e), (f) in our bilateral steerable pyramid, guided by the intensity image itself, detailed orientation are visible of textures, even nearby strong object edges.