Multi-view Image Fusion, Supplementary Material

Marc Comino Trinidad[1] Ricardo Martin Brualla[2] Florian Kainz[2] Janne Kontkanen[2]
[1]Polytechnic University of Catalonia, [2]Google

1. Contents

The supplementary material consists of this document and **Additional Results** in https://augmentedperception.github.io/pixelfusion/. These additional results contain high resolution images and comparisons for the applications proposed in the paper. The reader is strongly encouraged to view the images electronically in full-screen mode.

2. Comparison against HDR+

In this section we provide a brief comparison of our method and HDR+[2] using our Multi-view HDR fusion dataset. For this, we used the publicly available implementation [1] and we fed the algorithm both short and long exposure images. However, HDR+ expects a burst of similarly-exposed images, so we artificially reduced the exposure of the long-exposed image to match the short-exposed one. In Figure 1 we can see an example from our results. While our method is able to align and fuse both images, HDR+ struggles with the large disparity between them and produces alignment artifacts.

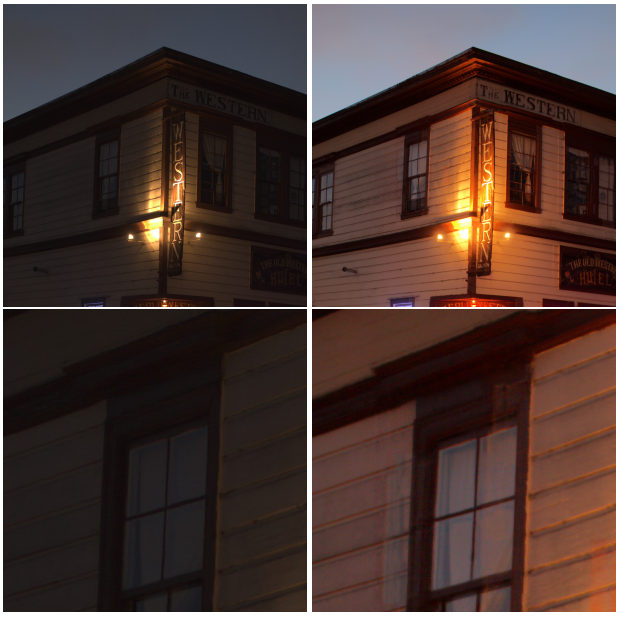

Figure 1: Left: our fused output. Right: HDR+ output. Please disregard the difference in overall color tone. This is because HDR+ implementation comes bundled with a tone mapping algorithm whereas we show our results without tone mapping, just gamma correction. The parallax between the original pair of images in this scene is of between 100 and 200 pixels. HDR+ fails to properly align both images and missalignment artifacts become apparent (check around the windows).

3. Failure Cases

In the following, we discuss typical failure cases of our algorithm and expand on the discussion on limitations in the paper.

3.1. Color transfer

As explained in the paper (Section 5, Limitations) we sometimes fail to properly align images in the presence of large disparities. Examples of this are shown in Figure 2 and in the **Additional Results** (see Section 1).

In Figure 2, first row, we can see that the area below the shirt is not correctly colorized. In the second row parts of the red scooter, right behind the dog are missing color. The third row exhibits various artifacts; the blue tint of the palm of the hand is perhaps the most distracting one.

Difficulties with large disparities can arise for several reasons:

1. Limited size of the search window. Traditionally optical flow algorithms do not search for correspondences within the entire image. Instead, to search for a correspondence for a pixel (x, y), the search is limited to a local window centered on (x, y). If the window is not large enough to cover the disparity, then the correspondence computation will fail.

Our search window is determined by the receptive field of the residual flow prediction network and the depth of the pyramid. As explained in the paper, we generally do not seem to be able to solve disparities that approach the search radius of the coarsest level (512). With the Yi Horizon camera we used, this happens when an object near the center of the frame is closer than about 14 cm.

The third row in Figure 2 shows areas with a disparity beyond this theoretical maximum, whereas the disparities in the other examples should fall within our search radius. This result might be improved if we used pyramids taller than nine levels, but we did not conduct that experiment.

2. Disocclusions. Sometimes the target image has areas that are not visible in the source image. While our method is quite efficient in inpainting these regions, occasionally the challenge is too difficult. This seems to be the case in Figure 2, second row. Here the orange color of the turn indicator and the red fender below it are not available in the source image, and the network does not attempt to guess them, but leaves these regions gray in the prediction.

3. Perspective distortion. When an object gets close to the camera, matching becomes more difficult because the two lenses see the object from different angles or scale. For an example of this, see the difference in scale of the thumb in the third row of Figure 2.

4. Insufficient training data. It is possible that our training data did not cover enough examples with large disparities. We have found that sharing weights across the pyramid helps with this problem and it seems theoretically possible that we could learn large image warps even from training data that does not contain them at all. However, we have not confirmed that this is actually the case.

To summarize, we were able to present plausible theories for the failures in the second and third row of Figure 2, but the first row needs further investigation.

3.2. Multi-frame HDR fusion

As shown in Figure 3, our method sometimes fails to align images or recover highlights in the Kalantari et al. [3] dataset. However, the method by Kalantari et al. has problems in the same example.

We hypothesize that there might be two different reasons for these problems:

1. The problem setup might be unnecessarily hard for the neural network. In this dataset, the network is tasked to produce the fused result in the mid-exposure frame. To produce correct predictions in areas that are saturated in the mid-exposure target frame, the network would have to find the corresponding pixels from the short exposure image. However, computing correspondence for saturated regions is prone to failure.

We suspect that the highlight recovery would be easier if the target view was the shortest exposure view instead. This way, the saturated regions would not need to be determined using the warped source image, but could be picked up directly from the target view. However, this would not come without cost, as aligning very dim pixels might become difficult instead.

2. The training set (74 images) might be too small to learn image warping. If this was the case, one could simply gather a larger dataset to address the problem. However, it might be beneficial to also consider transfer learning: i.e. pre-train the network for correspondence first and then fine-tune it for the specific task of HDR fusion. This way the problem-specific dataset could remain small.

References

[1] Hdr+ implementation. https://github.com/timothybrooks/hdr-plus. 1

[2] Samuel W Hasinoff, Dillon Sharlet, Ryan Geiss, Andrew Adams, Jonathan T Barron, Florian Kainz, Jiawen Chen, and Marc Levoy. Burst photography for high dynamic range and low-light imaging on mobile cameras. *ACM Transactions on Graphics (TOG)*, 35(6):192, 2016. 1

[3] Nima Khademi Kalantari and Ravi Ramamoorthi. Deep high dynamic range imaging of dynamic scenes. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2017)*, 36(4), 2017. 2, 3

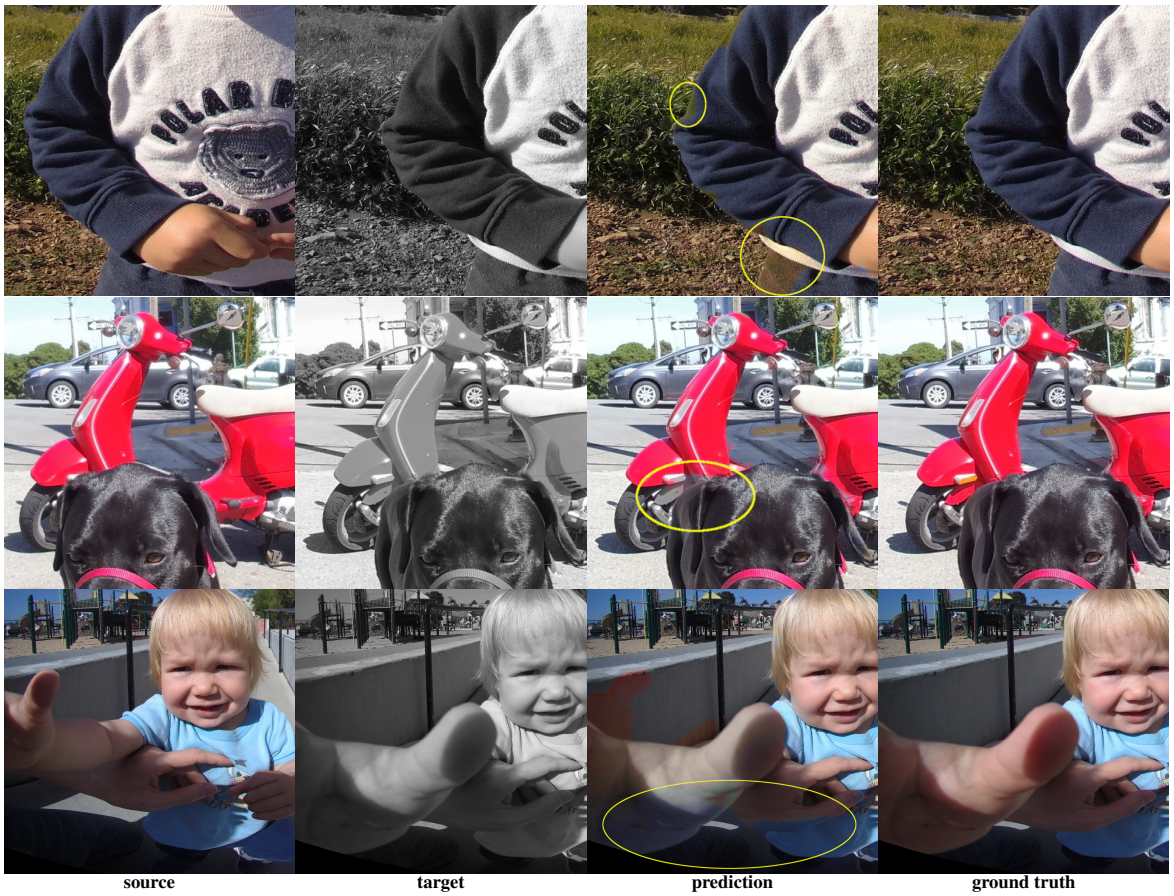

Figure 2: Color transfer failure cases in presence of large disparities. **First row**: the network has failed to correctly transfer color from the **source** to the **target** view. See the area below the shirt. **Second row**: This case demonstrates a large disocclusion that needs to be inpainted. The orange turn indicator and the red fender below it are not visible in the **source**, and the network produces a wrong **prediction** for that part of the image. **Third Row**: This example shows a disparity (approx. 880 pixels) that is beyond the theoretical maximum that our 9-level pyramid could resolve. The full images are provided in the **Additional Results** (see Section 1).

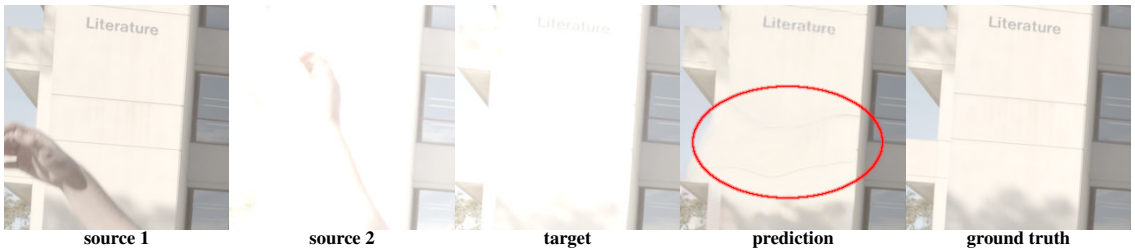

Figure 3: Multi-frame HDR fusion failure case. The network is asked to align **source 1** and **source 2** to the **target frame**, but because of the large saturated areas it fails to properly align the images and produces a **prediction** with artifacts typical for optical flow algorithms. The full images are shown in the **Additional Results** (see Section 1), where it is demonstrated that the method by Kalantari et al. [3] also has difficulties in this case.