Supplementary material for Accelerated Volume Rendering with Volume Guided Neural Denoising

S. Jabbireddy[1], S. Li[1], X. Meng[1], J. E. Terrill[2], A. Varshney[1]

[1]University of Maryland College Park, [2]National Institute of Standards and Technology, Maryland

1. Neural Bilateral Grid for denoising volumetric rendering

1.1. Bilateral Grid Filtering

A bilateral grid is a three-dimensional auxiliary data structure [CPD07] in image processing, introduced to speed up bilateral filtering. Bilateral filtering [AW95, SB97, TM98] performs a non-linear operation to smooth images while preserving edges. Bilateral grid-based denoising consists of three steps: splatting, filtering, and slicing. Given an input image I and a guide-image G (which could be the intensity image itself), the splatting operation projects the original two-dimensional spatial coordinates into a three-(or higher-) dimensional grid, where the first two dimensions (m,n) correspond to the spatial position in the input image, and the third dimension corresponds to the guide image intensity. Filtering or denoising is performed in this higher-dimensional space in the second step. Finally, slicing extracts a 2D image from the filtered grid using information from the guide image G. In practice, the resolution of the grid is much smaller than the original 2D image, mapping multiple pixels to the same grid cell.

1.2. Neural Bilateral Grid

The neural bilateral grid [MZV*20] extends the bilateral grid to denoise Monte Carlo ray-traced images with extremely low sample rates in real-time. We extend the neural bilateral grid for volumetric denoising. Specifically, we train the neural bilateral grid using our volumetric auxiliary features as input along with the noisy radiance values. The denoiser consists of a convolutional neural network to learn a guide image. The guide image creates a bilateral grid. In the next stage, spatial filtering is applied to the bilateral grid. Finally, the denoised image is obtained by slicing the filtered grid using the same learned guide image. Because of the small size of the bilateral grid, a light-weighted neural network is sufficient to learn the mapping, which facilitates real-time performance. Since previous research has established the advantage of incorporating multi-scale hierarchical networks, the denoiser employs bilateral grids at three different resolutions and computes the weighted average as the final denoised result.

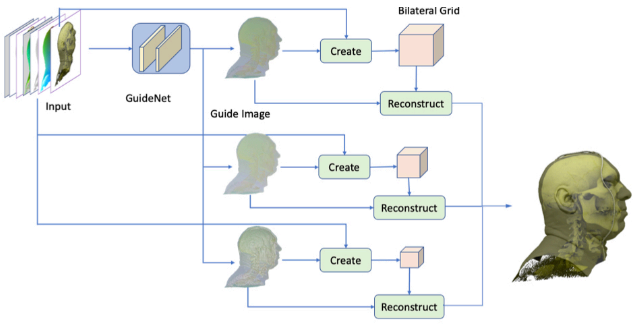

Figure 1: *The figure shows our volumetric denoising pipeline neural bilateral grid with three pyramid levels adapted from Meng* et al. *[MZV*20]. The GuideNet takes the noisy image and the auxiliary features as input and predicts the guide image, which guides bilateral filtering to produce the denoised image. The final output is a weighted combination of the denoised reconstructions from the three levels.*

2. Implementation details

2.1. Data

We validate the performance of our denoising method on a set of 3D volumes selected from https://klacansky.com/open-scivis-datasets and the National Library of Medicine – National Institutes of Health. The volumes correspond to the CT scans of *human head*, *stag beetle*, and *king snake*. Please refer to Table 1 for the details of these scenes.

Scene	Resolution	data type
Human Head	$512 \times 512 \times 294 = 77M\ voxels$	$16\ bits$
Stag Beetle	$832 \times 832 \times 494 = 341M\ voxels$	$16\ bits$
King Snake	$1024 \times 1024 \times 795 = 980M\ voxels$	$8\ bits$

Table 1: *Details of the scenes used in our experiments*

2.2. Training details

We train our volumetric denoiser using the images rendered from 100 different views around the volume and test on 50 novel views rendered similarly. We generate 50 patches of size 256×256 from each image and train for 200 epochs with a batch size of 128.

We consider three pyramid levels of bilateral grids in our denoising network. Each pyramid level corresponds to a particular bilateral grid resolution. In our experiments, we use the sampling factors as $(n_h, n_w, n_d) = \{(4,4,4), (8,8,8), (16,16,16)\}$ corresponding to the three pyramid levels. We generate guide images for each resolution and perform bilateral denoising independently. The final result is obtained as a weighted combination of the denoised images at the three levels, as shown in Figure 1.

The GuideNet consists of two dense convolutional layers with 20 hidden channels. We use leaky rectified linear unit (ReLU) activation for the first layer and sigmoid activation for the final layer. The tent filter used for filtering the bilateral grid is given by

$$T(x,y,z) = \max(1-|x|,0) \cdot \max(1-|y|,0) \cdot \max(1-|z|,0)$$

We use the mean absolute error between the denoised image and the ground truth image as the loss function.

References

[AW95] AURICH V., WEULE J.: Non-linear gaussian filters performing edge preserving diffusion. In *Mustererkennung 1995*. Springer, 1995, pp. 538–545. 1

[CPD07] CHEN J., PARIS S., DURAND F.: Real-time edge-aware image processing with the bilateral grid. *ACM Trans. Graph. 26*, 3 (jul 2007), 103–es. URL: https://doi.org/10.1145/1276377.1276506, doi:10.1145/1276377.1276506. 1

[MZV*20] MENG X., ZHENG Q., VARSHNEY A., SINGH G., ZWICKER M.: Real-time Monte Carlo Denoising with the Neural Bilateral Grid. In *Eurographics Symposium on Rendering - DL-only Track* (2020), Dachsbacher C., Pharr M., (Eds.), The Eurographics Association. doi:10.2312/sr.20201133. 1

[SB97] SMITH S. M., BRADY J. M.: SUSAN—a new approach to low level image processing. *International journal of computer vision 23*, 1 (1997), 45–78. 1

[TM98] TOMASI C., MANDUCHI R.: Bilateral filtering for gray and color images. In *Sixth international conference on computer vision (IEEE Cat. No. 98CH36271)* (1998), IEEE, pp. 839–846. 1