Direct Volume Rendering with Nonparametric Models of Uncertainty

Tushar Athawale, Bo Ma, Elham Sakhaee, Chris R. Johnson, *Fellow, IEEE,* and Alireza Entezari, *Senior Member, IEEE*

SUPPLEMENTARY MATERIAL

NONPARAMETRIC INTERPOLATION WITH BOX SPLINES

The key property of box splines is that their space is closed under convolution, and hence, an interpolation of histograms reduces to an analytical derivation of a higher order box spline with more direction vectors. For example, Fig. 1 is an illustration of the bilinear interpolation of random variables X_1, X_2, X_3, and X_4, where the PDF at each cell corner $\mathbf{v}_i$ is modeled using a histogram. We know that convolving four univariate B-splines (notice that each bin of a histogram is a scaled first-order B-spline) results in a cubic B-spline. Hence, the random variable at the interpolation point $\mathbf{v}$ assumes a distribution that constitutes the sum of translated cubic box splines. As Fig. 1 suggests, the resulting distribution can be multimodal or skewed, and is not restricted to follow any parametric form.

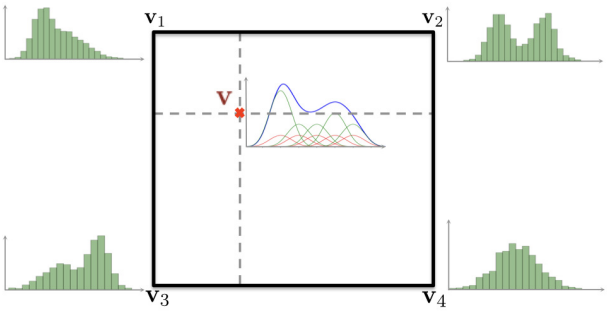

Fig. 1. Bilinear interpolation of nonparametric uncertain data. The distribution at an arbitrary point $\mathbf{v}$ is computed via a linear combination of cubic box splines, and can be skewed or multimodal.

Even though the box-spline method allows for interpolation of nonparametric distributions in a closed form [1], it is computationally expensive. For example, if n is the number of constant kernels at each cell vertex $\mathbf{v}_i$ in Fig. 1, then the number of cubic B-splines at the interpolated point $\mathbf{v}$ is n^4. In the case of 3D cells, the number of B-splines at the interpolated point is n^8. The exponential time complexity of the convolution-based nonparametric approach is a challenge to handle in the DVR reconstruction stage even when the system has high-performance hardware.

- *T. Athawale and C. R. Johnson are with the Scientific Computing & Imaging (SCI) Institute, University of Utah, Salt Lake City, UT, 84112. E-mails: {tushar.athawale,crj}@sci.utah.edu*
- *B. Ma, E. Sakhaee, and A. Entezari are with the Department of CISE at the University of Florida, Gainesville, FL, 32611. E-mails: {bbo, esakhaee, and entezari}@cise.ufl.edu*

Manuscript received April 19, 2005; revised August 26, 2015.

SAMPLE SIZE FOR QUANTILE INTERPOLATION

As we discussed in Sec. 5.1.1 of the paper, the reliability of quantile interpolation results is sensitive to the sample size or the number of ensemble members used for analysis. We describe our empirical approach to ensure sufficient sampling density for the dataset at hand and a user-chosen number of quantiles. Specifically, we perform qualitative and quantitative analysis of visualizations with varying sample sizes to empirically guide the selection of sampling density. We increase sample size in steps and keep track of how much visualizations vary upon increasing sample size. We stop when an increase in sample size does not change or affect the quality of visualizations. We now describe our qualitative and quantitative methods for empirically selecting sample sizes through demonstrations on the tangle (Fig. 3 of the paper) and Red Sea eddy simulation (Fig. 6 of the paper) datasets.

0.1 Qualitative Approach

In our qualitative method, we visually compare and contrast the quality of visualizations for incremental sample sizes. Fig. 2 visualizes such a result for the Red Sea eddy simulation dataset. In Fig. 2, the visualizations appear similar for the sample sizes $n = 20, 30, 40$. On the contrary, a few reconstruction artifacts can be seen for the sample size $n = 10$, indicated by the dotted white boxes. Thus, a sufficient sampling density can be estimated for the dataset through a visual comparison of results for varying sample sizes. Fig. 3 visualizes the result for the tangle dataset with varying sample sizes. Although the visualizations look similar overall for all sample sizes, i.e., $n = 5, \cdots, 50$, the visualizations in Fig. 3(a)-(c) display a relatively higher instability of reconstructions compared to Fig. 3(d)-(f).

0.2 Quantitative Approach

In our quantitative method, we compute the average structural similarity index (SSIM) [2] between the RGB channels

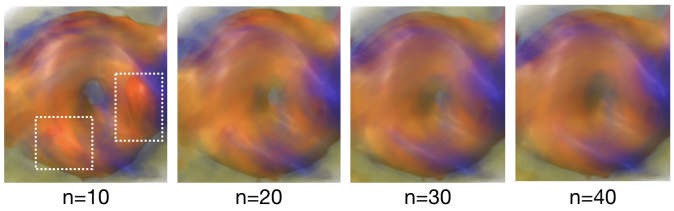

Fig. 2. Red Sea eddy simulations: Effect of sample size on visualizations with the quantile mean integration (eight quantiles). White boxes illustrate the visualization artifacts caused by insufficient sampling density.

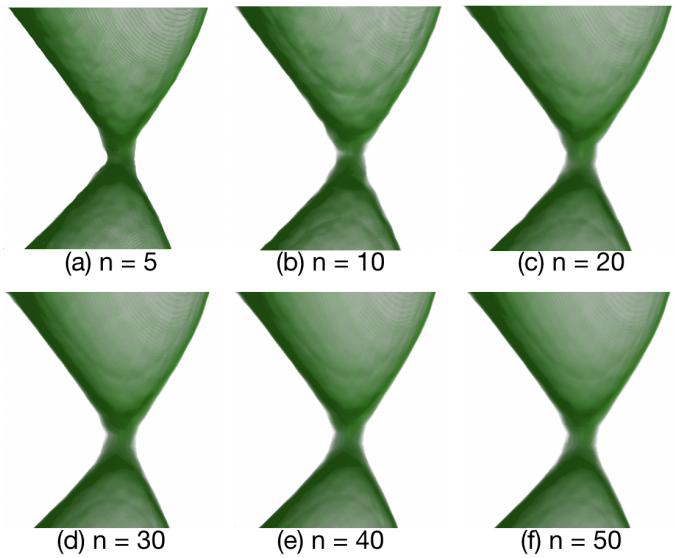

Fig. 3. Tangle dataset: Effect of sample size on visualizations with the quantile mean integration (eight quantiles).

of visualizations for a pair of datasets, in which the sample sizes for the datasets differ by five samples. We then increase the sample size and study its effects on SSIM. We stop increasing the sample size when high SSIM values are consistently observed for subsequent pairs of datasets. High SSIM values indicate the high structural similarity of visualizations, i.e., visualizations do not change in terms of quality with an increase in the number of samples.

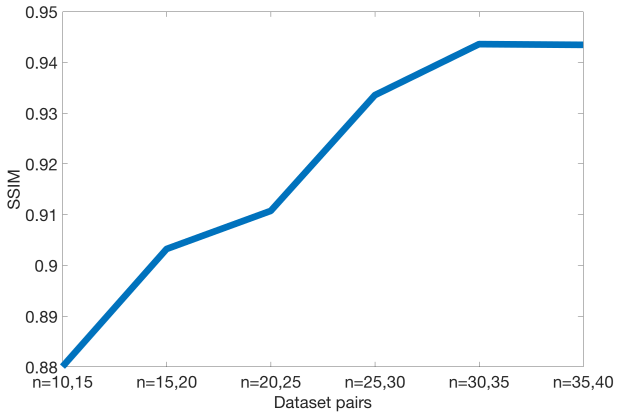

Fig. 4. Effect of sample size on the SSIM for visualizations of Red Sea eddy simulations using the quantile mean integration (eight quantiles).

Fig. 4 visualizes a result for our quantitative method on the Red Sea eddy simulation dataset. The horizontal axis in Fig. 4 represents the sample size pairs, where the sample size for each pair differs by five samples. The vertical axis plots the SSIM computed for visualizations produced using each pair. With an increase in the sample size, the SSIM increases, thus indicating higher similarity of visualizations for higher sampling density. The SSIM appears consistently greater than 0.9 for sample size $n >= 20$. Thus, $n = 20$ may be regarded as a sufficient sampling density for generating visualizations. However, higher sampling will always result in improved reliability of reconstruction. Fig. 5 visualizes a plot indicating the effect of sample size on the SSIM for the tangle dataset. In Fig. 5, the SSIM is consistently greater than 0.9 for the sample size $n >= 35$. Thus, $n = 35$ may be regarded as a sufficient sampling density for generating the tangle dataset visualizations.

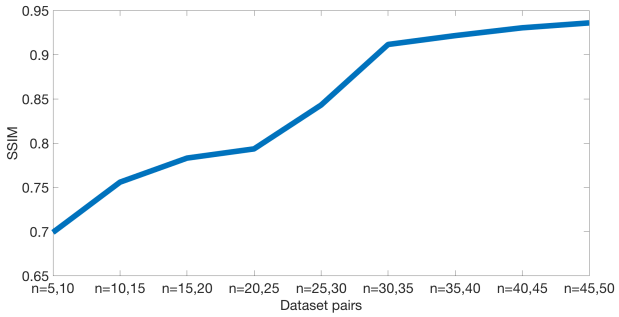

Fig. 5. Effect of sample size on the SSIM for visualizations of the tangle dataset using the quantile mean integration (eight quantiles).

EVEN VS. UNEVEN QUANTILE VALUES

The approach described in Sec. 3 of the manuscript assumes a fixed quantile value $qval$ for quantile interpolation. When quantile values are not fixed, there are two possible scenarios. First, quantile values at all grid vertices are different. Second, the quantile values are uneven at a single grid vertex, but the same uneven quantile value combination is repeated at each grid vertex. We briefly discuss these two scenarios.

Scenario 1: We discuss the case when quantile values are different at all grid vertices. We revisit the 1D quantile interpolation formula in Eq. 1 of the paper. Let $qval_1$ denote a quantile value for the first quantile Q_{11} (ordered by CDF) of a random variable X_1. Similarly, let $qval_2$ denote a quantile value for the first quantile Q_{21} of a random variable X_2, where $qval_2 \neq qval_1$. Let w_1 and w_2 denote the widths of the quantiles Q_{11} and Q_{21}, respectively. We, therefore, get the quantile probability densities as: $Pr(Q_{11}) = \frac{qval_1}{w_1}$ and $Pr(Q_{21}) = \frac{qval_2}{w_2}$ (see Sec. 3 in the paper). Applying a 1D quantile interpolation formula in Eq. 1 of the paper to the quantile probability densities, we get:

$$Pr(Q_1) = \frac{(qval_1)(qval_2)}{(qval_1)\alpha w_{2j} + (qval_2)(1-\alpha)w_{1j}} \quad (1)$$

Since $qval_1 \neq qval_2$, the product cumulative density, i.e., $qval_1 \times qval_2$ is not distributed over interpolated quantile

widths, i.e., $\alpha w_{2j} + (1-\alpha)w_{1j}$, as we observe in Eq. 1 of the paper. It is, however, spread over an interval weighted by the quantile values, i.e., $(qval_1)\alpha w_{2j} + (qval_2)(1-\alpha)w_{1j}$. This scenario is, thus, nontrivial and we plan to address it in our future work.

Scenario 2: Consider a two-quantile case for 1D quantile interpolation. Suppose we fix uneven quantile values at each grid vertex as $[0.8, 0.2]$. In such a case, Eq. 1 boils down to Eq. 1 in the paper since the j'th quantiles for each random variable, i.e., Q_{ij} hold the same quantile value, i.e., $qval_1 = qval_2$ in Eq. 1. The uneven quantile value assumption in such cases can significantly improve DVR reconstruction accuracy. We demonstrate improved reconstruction accuracy for a sample dataset in Fig. 6 for fixed uneven quantile value combination assumption. The quantile interpolation results presented in Fig. 3 of the paper use even quantile values at each grid vertex. For example, in Fig. 3m of the paper, the quantile value combination is even, i.e., [0.5,0.5] for $qval = 0.5$ at each grid vertex. In Fig. 6, we present a result for uneven quantile values at a grid vertex for the same tangle dataset used in the paper. Uneven quantile value combination can be discovered using a clustering algorithm such as k-means. At each vertex we run the k-means algorithm with $k = 2$ for clustering noise samples, in which these noise samples are ordered or sorted. We then find a cumulative density for each k-means cluster using a kernel density PDF estimation from noise samples. Because of the way we injected bimodal noise distribution at each vertex for the tangle dataset (see paragraph 1 of Sec. 5.1 of the paper), we get the quantile value combinations close to $[0.8, 0.2]$ at every grid vertex using the k-means clustring. We then take the average of all quantile-value combinations and set the uneven quantile-value combination at each grid vertex as $[0.78, 0.22]$. Quantile interpolation visualization using uneven quantile values improves reconstruction accuracy remarkably well, as can be seen in Fig. 6b, with RMSE equal to 0.0002.

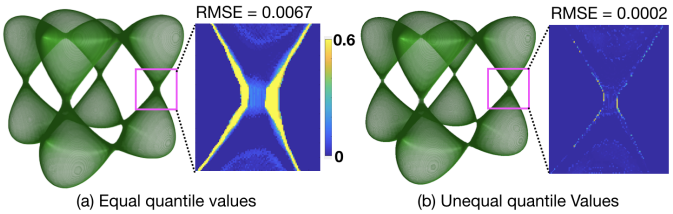

Fig. 6. Even (image (a)) vs. uneven (image (b)) quantile values for DVR of the tangle dataset using quantile mean technique. Unequal quantile value assumption (Scenario 2) exceptionally improves DVR reconstructions.

However, it is again not clear how the fixed uneven quantile value combination assumption can be useful in practice. Hence, we again leave the investigation of this scenario as our future work.

REFERENCES

[1] E. Sakhaee and A. Entezari, "A statistical direct volume rendering framework for visualization of uncertain data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 12, pp. 2509–2520, Dec 2017.

[2] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.