Bessel-NeRF: An Analytic Integrated Positional Encoding for Neural Radiance Fields

Brian K. S. Isaac-Medina
Department of Computer Science, Durham University, UK
brian.k.isaac-medina@durham.ac.uk

1. Introduction

Neural Radiance Fields (NeRF) [4] is an image-based rendering technique that has attracted significant attention because of its ability to synthesize novel views with good performance. NeRFs use a multi-layer perceptron to predict the colour and density of points in space (*i.e.*, the field of colour and densities defined by the scene), sampling them from the rays with origin at the centre of a viewing camera, and that pass through the pixels of the image formed by that camera. These points are encoded using a high-frequency function, called positional encoding, which avoids the natural bias of NeRFs towards predicting low-frequency functions [4, 5]. The predicted properties of these points are then combined using alpha compositing to render an image. However, the composition of points along a ray may cause ambiguous representations that lead to further artifacts such as aliasing. A recent variant, mip-NeRF [1], proposes the casting of cones instead of rays in order to create volumetric representations of a region of the scene. This parameterization involves summarizing the information of a cone frustum using the integral of a positional encoding, which in turn has no analytical solution. In order to deal with this, mip-NeRF approximates this integral as the expected value of a multi-variate Gaussian distribution parameterized by the geometric properties of the frustum, which they call Integrated Positional Encoding (IPE). This approximation is reliable for nearly isotropic frustums but degrades with highly elongated regions, which arise when dealing with distant objects. For this reason, in this extended abstract, we detail an alternative solution to the IPE by going back to the task of calculating the integral of the positional encoding over the cone frustums. While this approach requires the computation of an integral with no analytic solution, we approximate it by using a linear expansion of the integrand around the midpoint of the cone frustum. This involves the use of Bessel functions of the first kind, which is the motivation to call our method *Bessel-NeRF*, Our proposed approach, along with a concurrent work [3], is one of the first attempts to tackle the computation of this integral and may lead to further improvements for more challenging tasks.

2. Preliminaries

In this section, the formulation of NeRF [4] and mip-NeRf [1] are presented, which form the basis of our method.

2.1. Neural Radiance Fields

A NeRF learns an implicit representation of a 3D scene from a set of 2D input images, taking as input a point $\mathbf{x} \in \mathbb{R}^3$ and a viewing direction $\hat{\mathbf{d}} \in S_2$, where S_2 is the unit sphere, and getting the corresponding colour $\mathbf{c} \in \mathbb{R}^3$ and density $\sigma \in [0, +\infty)$ through a neural network f with parameters Θ, such that

$$(\mathbf{c}, \sigma) = f(\mathbf{x}, \hat{\mathbf{d}}; \Theta)\,. \tag{1}$$

In order to induce high-frequency features, a positional encoding $\gamma : \mathbb{R} \to \mathbb{R}^{2L}$ acting as a map to a higher dimensional space with high-frequency functions is used [4,5]. In particular, NeRF uses the positional encoding

$$\begin{aligned}\gamma(\mathbf{x}) = \big[&\sin(2^0\mathbf{x}), \cos(2^0\mathbf{x}), \dots, \\ &\sin(2^{L-1}\mathbf{x}), \cos(2^{L-1}\mathbf{x})\big]^\top,\end{aligned} \tag{2}$$

where γ is applied to each coordinate of $\mathbf{x}$ and each component of $\hat{\mathbf{d}}$ independently.

NeRF uses a sampling strategy consisting in taking random points along the rays that pass through the pixels of each image. This ray is represented by $\mathbf{r}(t) = t\mathbf{d} + \mathbf{o}$, where $\mathbf{o}$ is the camera centre position and $\mathbf{d}$ is the vector that goes from $\mathbf{o}$ to the pixel in the image plane. The ray is divided into N intervals and the points $\mathbf{r}(t_i)$ are drawn from a uniform distribution over each interval, such that

$$t_i \sim \mathcal{U}\left[t_n + \frac{i-1}{N}(t_f - t_n), t_N + \frac{i}{N}(t_f - t_n)\right]\,, \tag{3}$$

where t_n and t_f are the near and far planes. In this sense, the colour and density of each point over the ray is obtained by $(\mathbf{c}_i, \sigma_i) = f(\gamma(\mathbf{r}(t_i)), \gamma(\mathbf{d}/\|\mathbf{d}\|); \Theta)$.

Finally, the pixel colour $\hat{C}(\mathbf{r})$ is obtained using numerical quadrature

$$\hat{C}(\mathbf{r}) = \sum_{i=1}^{N} T_i(1 - \exp(-\sigma_i\delta_i)) , \tag{4}$$

$$T_i = \exp\left(-\sum_{j=1}^{i-1} \sigma_j\delta_j\right) , \tag{5}$$

where $\delta_i = t_{i+1} - t_i$. The sampling is carried out hierarchically by using coarse $\hat{C}_c$ and fine $\hat{C}_f$ samplings, where the 3D points in the latter are drawn from the PDF formed by the weights of the density values of the coarse sampling. The loss is then the combination of the mean-squared error of the coarse and fine renderings for all rays $\mathbf{r} \in \mathcal{R}$, *i.e.*,

$$\mathcal{L} = \sum_{\mathbf{r}\in\mathcal{R}} \left[\|\hat{C}_c(\mathbf{r}) - C(\mathbf{r})\|_2^2 + \|\hat{C}_f(\mathbf{r}) - C(\mathbf{r})\|_2^2\right] . \tag{6}$$

2.2. Mip-NeRF

Barron *et al*. [1] noticed that the use of a ray per pixel may lead to some artifacts (*e.g*., aliasing and blurring) that are caused by ambiguous point features at different views or scales. They proposed mip-NeRF, which is similar to NeRF but it uses cone tracing instead of rays. This change has the direct consequence of replacing ray intervals by conical frustums $F(\mathbf{d}, \mathbf{o}, \dot{\rho}, t_i, t_{i+1})$, where $\dot{\rho}$ is the radius of the circular section of the cone at the image plane. This leads to the need for a new positional encoding that summarizes the function in Eq. (2) over the region defined by the frustum. The proposed formulation, called integrated positional encoding (IPE), is given by

$$\gamma_I(\mathbf{d}, \mathbf{o}, \dot{\rho}, t_i, t_{i+1}) = \frac{\iiint_F \gamma(\mathbf{x})dV}{\iiint_F dV} . \tag{7}$$

Since the integral in the numerator of Eq. (7) has no closed-form solution, mip-NeRF proposes to approximate it by considering the conic frustums as multivariate Gaussians, obtaining the relation

$$\begin{aligned}\gamma^*(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \mathbb{E}_{x\sim\mathcal{N}(\mathbf{P}\boldsymbol{\mu},\mathbf{P}\boldsymbol{\Sigma}\mathbf{P}^\top)}\left[\gamma(\mathbf{x})\right] \\ &= \begin{bmatrix}\sin(\mathbf{P}\boldsymbol{\mu}) \circ \exp(-(1/2)\mathrm{diag}(\mathbf{P}\boldsymbol{\Sigma}\mathbf{P}^\top)) \\ \cos(\mathbf{P}\boldsymbol{\mu}) \circ \exp(-(1/2)\mathrm{diag}(\mathbf{P}\boldsymbol{\Sigma}\mathbf{P}^\top))\end{bmatrix} ,\end{aligned} \tag{8}$$

where $\boldsymbol{\mu} = \mathbf{o} + \mu_t\mathbf{d}$ is the centre of the Gaussian for a frustum defined by $\mathbf{o}$ and $\mathbf{d}$ with mean distance along the ray μ_t, $\boldsymbol{\Sigma}$ is the covariance matrix, $\circ$ denotes element-wise product and

$$\mathbf{P} = \begin{bmatrix}1 & 0 & 0 & 2 & 0 & 0 & & 2^{L-1} & 0 & 0 \\ 0 & 1 & 0 & 0 & 2 & 0 & \dots & 0 & 2^{L-1} & 0 \\ 0 & 0 & 1 & 0 & 0 & 2 & & 0 & 0 & 2^{L-1}\end{bmatrix}^\top . \tag{9}$$

This formulation proved to be accurate for bounded scenes where a central object is the main part of the scene and no background information is present. However, the approximation gets worse for highly elongated frustums, as noted by Barron *et al*. [2].

3. Bessel-NeRF

We now turn our attention to the solution of Eq. (7). First, we consider a coordinate system $[x', y', z']^\top$ with the z' axis aligned to $\mathbf{d}$ and a rotation matrix R, such that its transform to the world coordinate system is given by

$$[x, y, z]^\top = \zeta(x', y', z') = \mathrm{R}[x', y', z']^\top + \mathbf{o} . \tag{10}$$

Similar to mip-NeRF [1], we use an axis-aligned cone parameterization $[x', y', z']^\top = \phi(\rho, t, \theta) = [\rho t\cos\theta, \rho t\sin\theta, t]^\top$ for $\theta \in [0, 2\pi)$, $t \geq 0$ and $0 \leq \rho \leq \dot{\rho}$ to have our final transform to the Cartesian space

$$\begin{aligned}[x, y, z]^\top &= \xi(\rho, t, \theta) = (\zeta \circ \phi)(\rho, t, \theta) \\ &= \mathrm{R}[\rho t\cos\theta, \rho t\sin\theta, t]^\top + \mathbf{o} .\end{aligned} \tag{11}$$

Then, we obtain the volumen of the cone frustum by solving the integral in the denominator of Eq. (7), *i.e.*,

$$\begin{aligned}\iiint_F dV &= \int_{t_i}^{t_{i+1}} \int_0^{\dot{\rho}} \int_{-\pi}^{\pi} \det(\mathbf{J}_\xi)d\theta d\rho dt \\ &= \int_{t_i}^{t_{i+1}} \int_0^{\dot{\rho}} \int_{-\pi}^{\pi} \det(\mathrm{R}\mathbf{J}_\phi)d\theta d\rho dt \\ &= \int_{t_i}^{t_{i+1}} \int_0^{\dot{\rho}} \int_{-\pi}^{\pi} \det(\mathrm{R})\det(\mathbf{J}_\phi)d\theta d\rho dt \\ &= \int_{t_i}^{t_{i+1}} \int_0^{\dot{\rho}} \int_{-\pi}^{\pi} \rho t^2 d\theta d\rho dt \\ &= \frac{1}{3}\pi\dot{\rho}^2(t_{i+1}^3 - t_i^3) ,\end{aligned} \tag{12}$$

where $|\mathbf{J}_f|$ is the determinant of the Jacobian matrix of f and the determinant of R is 1 since it is a rotation matrix.

In order to solve the integral in the numerator of Eq. (7), we use the same parameterization in Eq. (11) and consider the sine function $\sin(2^l\mathbf{x})$ of the positional encoding in Eq. (2) for the l-th frequency. Then, the numerator of the IPE for the k-th coordinate is calculated as

$$\begin{aligned}&\iiint_F \sin(2^l\mathbf{x}_k)dV \\ &\quad= \int_{t_i}^{t_{i+1}} \int_0^{\dot{\rho}} \int_{-\pi}^{\pi} \sin(2^l\xi(\rho, t, \theta)_k)\rho t^2 d\theta d\rho dt .\end{aligned} \tag{13}$$

Now we proceed to solve the IPE for the k-th coordinate. Substituting and expanding Eq. (11) in Eq. (13) and having r_{ij} as the elements of the rotation matrix R, we obtain

$$\gamma_I(\mathrm{R}, \mathbf{o}, \dot{\rho}, t_i, t_{i+1})_k = \int_{t_i}^{t_{i+1}} \int_0^{\dot{\rho}} \int_{-\pi}^{\pi} \sin(2^l(r_{k1}\rho t \cos\theta + r_{k2}\rho t \sin\theta + r_{k3}t + o_k))\rho t^2 d\theta d\rho dt \,. \tag{14}$$

Then, by using the trigonometric identity $a\cos\theta + b\sin\theta = \sqrt{a^2+b^2}\cos(\theta + \arctan(-b/a))$, we can write the Eq. (14) as

$$\begin{aligned}\gamma_{I\,k} &= \int_{t_i}^{t_{i+1}} \int_0^{\dot{\rho}} \int_{-\pi}^{\pi} \sin(A\cos(\theta + B) + C)\rho t^2 d\theta d\rho dt & (15)\\ &= \int_{t_i}^{t_{i+1}} \int_0^{\dot{\rho}} \int_{-\pi}^{\pi} \sin(A\cos\theta + C)\rho t^2 d\theta d\rho dt & (16)\\ &= \int_{t_i}^{t_{i+1}} \int_0^{\dot{\rho}} \int_{-\pi}^{\pi} \sin(A\cos\theta)\cos(C)\rho t^2 d\theta d\rho dt \\ &+ \int_{t_i}^{t_{i+1}} \int_0^{\dot{\rho}} \int_{-\pi}^{\pi} \cos(A\cos\theta)\sin(C)\rho t^2 d\theta d\rho dt\,, & (17)\end{aligned}$$

where

$$A = 2^l \rho t \sqrt{r_{k1}^2 + r_{k2}^2} \tag{18}$$
$$B = \arctan(-r_{k2}/r_{k1}) \tag{19}$$
$$C = 2^l(r_{k3}t + o_k)\,, \tag{20}$$

and in Eq. (16) we used the fact that the integral over a full period of $\cos(\theta+\varepsilon)$ is the same for any value of ε. We omit the arguments of γ_I for clarity. It is noted that $\int_0^{\pi/2} \sin(A\cos\theta)d\theta = -\int_{\pi/2}^{\pi} \sin(A\cos\theta)d\theta$, and since $\sin(A\cos\theta)$ is even, the first part of Eq. (17) is 0. Given that the second part of Eq. (17) is also even, we can simplify it to

$$\gamma_{I\,k} = 2\int_{t_i}^{t_{i+1}} \int_0^{\dot{\rho}} \int_0^{\pi} \cos(A\cos\theta)\sin(C)\rho t^2 d\theta d\rho dt \,. \tag{21}$$

In order to solve Eq. (21), we need to use the Bessel function of first kind $J_\nu(x)$, which can be defined in its integral form as

$$J_\nu(x) = \frac{1}{\pi}\int_0^{\pi} \cos(\nu\tau - x\cos\tau)d\tau \,. \tag{22}$$

Using this relation, Eq. (21) is reduced to

$$\gamma_{I\,k} = 2\pi \int_{t_i}^{t_{i+1}} \int_0^{\dot{\rho}} J_0(A)\sin(C)\rho t^2 d\rho dt \,. \tag{23}$$

We now rewrite Eq. (23) as

$$\gamma_{I\,k} = 2\pi \int_{t_i}^{t_{i+1}} \int_0^{\dot{\rho}} J_0(A'\rho)\sin(C)\rho t^2 d\rho dt \tag{24}$$
$$A' = A/\rho = 2^l t\sqrt{r_{k1}^2 + r_{k2}^2}\,. \tag{25}$$

The integral in Eq. (24) can be solved using the relation $\int_0^a J_0(bu)udu = (a/b)J_1(ab)$. Then

$$\gamma_{I\,k} = 2\pi\dot{\rho}\int_{t_i}^{t_{i+1}} \frac{J_1(A'\dot{\rho})}{A'}\sin(C)t^2 dt \,. \tag{26}$$

Substituting Eqs. (20) and (25) in Eq. (26)

$$\gamma_{I\,k} = \lambda_k \int_{t_i}^{t_{i+1}} J_1(\alpha_k t)\sin(\beta_k t + \psi_k)t dt \tag{27}$$
$$\lambda_k = \frac{2\pi\dot{\rho}}{2^l\sqrt{r_{k1}^2 + r_{k2}^2}} \tag{28}$$
$$\alpha_k = 2^l \dot{\rho}\sqrt{r_{k1}^2 + r_{k2}^2} \tag{29}$$
$$\beta_k = 2^l r_{k3} \tag{30}$$
$$\psi_k = 2^l o_k\,. \tag{31}$$

The integral in Eq. (27) has no closed-form solution. Therefore, we linearly approximate the integrand $f(t) = J_1(\alpha_k t)\sin(\beta_k t + \psi_k)t$ around the midpoint $\mu_i = (t_i + t_{i+1})/2$

$$f(t) \approx J_1(\alpha_k\mu_i)\sin(\beta_k\mu_i + \psi_k)\mu_i + \frac{df}{dt}(\mu_i)(t - \mu_i)\,. \tag{32}$$

We can rewrite the integration limits of Eq. (27) as $t_i = \mu_i - \delta_i/2$ and $t_{i+1} = \mu_i + \delta_i/2$. By integrating Eq. (32) with these integration limits, it can be seen that the linear term $\frac{df}{dt}(\mu_i)(t - \mu_i)$ vanishes. Our approximation of the IPE for the k-th coordinate is then

$$\gamma^*_{I\,k} = \lambda_k J_1(\alpha_k\mu_i)\sin(\beta_k\mu_i + \psi_k)\mu_i\delta_i\,. \tag{33}$$

We call our approximation Bessel Integrated Positional Encoding (BIPE) and forms the basis of Bessel-NeRF. Similarly, the BIPE for the cosine function is

$$\gamma^*_{I\,k} = \lambda_k J_1(\alpha_k\mu_i)\cos(\beta_k\mu_i + \psi_k)\mu_i\delta_i\,. \tag{34}$$

Finally, we define the vectors

$$\boldsymbol{\eta} = \mathbf{P}\left(\mathbf{R}_{:,1}^{\circ 2} + \mathbf{R}_{:,2}^{\circ 2}\right)^{\circ 1/2} \tag{35}$$

$$\boldsymbol{\lambda} = 2\pi\dot{\rho}(\boldsymbol{\eta})^{\circ -1} \tag{36}$$

$$\boldsymbol{\alpha} = \dot{\rho}\boldsymbol{\eta} \tag{37}$$

$$\boldsymbol{\beta} = \mathbf{R}_{:,3}\,, \tag{38}$$

where $\mathbf{R}_{:,j}$ is the j-th column of $\mathbf{R}$ and $(\cdot)^{\circ n}$ refers to the element-wise n-th power, to combine the Eqs. (33) and (34) into our final BIPE feature

$$\boldsymbol{\gamma}_I^* = \mu_i \delta_i \begin{bmatrix} \boldsymbol{\lambda} \circ J_1\left(\mu_i \boldsymbol{\alpha}\right) \circ \sin(\mathbf{P}(\mu_i \boldsymbol{\beta} + \mathbf{o})) \\ \boldsymbol{\lambda} \circ J_1\left(\mu_i \boldsymbol{\alpha}\right) \circ \cos(\mathbf{P}(\mu_i \boldsymbol{\beta} + \mathbf{o})) \end{bmatrix}. \tag{39}$$

4. Conclusion and Future Work

In this extended abstract, we present Bessel-NeRF, an alternative parameterisation of mip-NeRF that goes back to the task of integrating the positional encoding over a conical frustum. This is a work in progress and the results presented are the outcomes of the theoretical formulation of a new positional encoding. The next steps in this work are to implement Bessel-NeRF and compare it against mip-NeRF and other formulations.

References

[1] Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5855–5864, 2021. 1, 2

[2] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5470–5479, 2022. 2

[3] Brian K. S. Isaac-Medina, Chris G. Willcocks, and Toby P. Breckon. Exact-NeRF: An exploration of a precise volumetric parameterization for neural radiance fields. *CVPR*, 2023. 1

[4] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. 1

[5] Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred A. Hamprecht, Yoshua Bengio, and Aaron Courville. On the spectral bias of neural networks. 2018. 1