Deformable Models and Geodesic Methods for image analysis

Review of *Neural Density-Distance Fields*

Gauthier MULTARI
ENS Paris-Saclay
gauthier.multari@ens-paris-saclay.fr

March 31, 2024

This report attempts to review the paper *Neural Density-Distance Fields*, Ueda et al. [9]

1 Synthetic overview

1.1 Problem

Historically, the most commonly used methods for volume representation aim to discretize either the density or the distance from the surface at each position into voxels. However, by construction, voxels lead to a data complexity cubic in resolution. The higher the spatial resolution, the harder it is to handle the amount of data required. Alternatives to such representations, with attempts to represent this information in grid-like structures, struggle to handle parts of the geometry with frequencies with higher frequencies than the Nyquist frequency.

With the rise of deep learning, a new approach has been proposed: representing continuous signals by using neural networks to encode implicit surfaces. The methods based on this representation are called neural fields [4]. This is extremely powerful because, given that they have enough parameters, neural networks can encode continuous signals over arbitrary dimensions. Furthermore, as the networks are adaptive, they can modulate the information they hold at a given frequency, so they can yield more information in high frequency regions for instance. Moreover, the use of a continuous representation allows modifications of both the input and output dimensions without modifiying the model itself. On top of that, such representation of the data requires significantly less parameters than historical methods for a given resolution. Finally, the neural fields can be seen as a continuously differentiable field when using a smooth activation function in the neural networks. The presented work starts from this realization and aims to constrain both the distance and the density fields to improve the reconstruction quality.

1.2 Equations and numerical methods

Here we describe the main equations of the paper, details on the calculations and the practical aspects can be found in the paper summary (2).

Let the distance field $D_b(\mathbf{p})$ that describes the distance to the nearest surface for a location $\mathbf{p} \in \mathbb{R}^3$ and the depth value $d_b(\mathbf{p}, \mathbf{v})$ over the viewing direction $\mathbf{v} \in \mathcal{S}^2$. We can interpret $D_b(\mathbf{p})$ as the minimum of the depth: $D_b(\mathbf{p}) := \min_{\mathbf{v}\in S^2} [d_b(\mathbf{p}, \mathbf{v})]$.

In order to be able to render scenes with varying density distributions, the authors present equation 1. For a point on the ray $\mathbf{r}(t) = \mathbf{p} + t\mathbf{v}$ with the visible range $[t_n, t_f]$, the color of each ray $\mathbf{C}(\mathbf{r})$ is obtained through the integral of each color $\mathbf{c}(\mathbf{r}(t), \mathbf{v})$ multiplied by transmission rate $T(t) = \exp\left(-\int_{t_n}^{t} \sigma(\mathbf{r}(s))ds\right)$:

$$C(\mathbf{r}) := \int_{t_n}^{t_f} T(t)\sigma(\mathbf{r}(t))\mathbf{c}(\mathbf{r}(t), \mathbf{v})dt \tag{1}$$

The depth $d(\mathbf{p}, \mathbf{v})$ is defined to be an integral of the depths at each point:

$$d(\mathbf{p}, \mathbf{v}) := t_n + \int_{t_n}^{t_f} tT(t)\sigma(\mathbf{r}(t))dt \tag{2}$$

But, given that calculating the depths over all directions is extremely expensive computationally, the distance field is defined independently of the viewing direction: $D(\mathbf{p}) := \min_{\mathbf{v} \in S^2}[d(\mathbf{p}, \mathbf{v})]$.

Starting from this basis, the authors derive an expression that converts the distance and its gradient into density using the fact that the distance field is an integral of a polynomial about density:

$$\sigma(\mathbf{p}) \simeq \frac{1 - \|\nabla D(\mathbf{p})\|_2}{D(\mathbf{p})} \tag{3}$$

By using this expression, a density field consistent with the distance field represented by the neural field can be obtained in a differentiable form. This means that it is possible to learn mutually constrained distance and density fields.

1.3 Links with the course

This article is not directly linked to the methods seen in this class. However, they have a similar goal as adaptive sampling in that both these fields aim to represent data in a computationnaly efficient manner by reducing the information in low frequency regions and increasing it in high frequency regions. The difference is that neural fields provide a continuous, function-based approach, where point sampling offers a discrete approximation. Furthermore, as we have seen, geodesic and minimal path techniques can be used for 3D reconstruction and scene understanding, and the approach of reciprocally contraining density and distance fields could allow for the use of these methods to improve the reconstruction results.

1.4 Novelty of the method and main results

This method applies a reciprocal constraint between distance and density fields, making use of both their advantages. Unlike traditional methods that might focus on one or the other, NeDDF effectively captures the essence of objects and phenomena without clear boundaries, like fur or smoke, which has been a challenge in 3D modeling.

One of the key innovations of NeDDF is its ability to ensure consistency between the distance and density fields. This is crucial for accurate localization and high-quality registration, allowing the model to understand and predict the 3D structure of a scene from sparse data points effectively. This consistency leads to robustness against initial value settings and achieves fast convergence, which is invaluable for real-world applications where data may be incomplete or noisy.

Moreover, despite its focus on spatial understanding, NeDDF does not sacrifice visual quality. It achieves results comparable to NeRFs in generating novel views of a scene, indicating that NeDDF can provide both detailed visual synthesis and accurate spatial understanding. This dual capability addresses a common trade-off in previous methods and opens new possibilities for capturing and rendering complex materials and phenomena within virtual environments.

1.5 Weaknesses

This approach, while innovative retains the main drawbacks of neural fields approches. We can cite high computational demand, especially when processing complex scenes or requiring high-resolution outputs. The ability of NeDDF to generalize across vastly different scenes or object types without retraining or adjustments remains an area is not something discussed by the authors. While NeDDF shows promising results in static environments, its effectiveness in dynamic scenes, where objects move or change over time, may require further exploration and adaptation to ensure accuracy and reliability. Additionally, neural field methods typically rely on substantial amounts of training data to achieve high fidelity and accuracy. Therefore, the efficiency of NeDDF in scenarios with sparse or limited data could be another weakness.

2 Summary of the article

2.1 Previous work

Before explaining the proposition of the authors, it is necessary to detail their intuitions regarding the two main objects used for their work, the Density Field and the Distance Field.

Density Field Given a 3D position as the input, the density field returns its corresponding volume density. The authors describe the density field as being characterized by high expressiveness, meaning that it holds a great power of representation. For instance, a low value of the field could be the representation of an altered light transmission, such as in glass or smoke. For a more complex scene, we can retrieve the reconstruction information at a given point by considering the light interaction as a function of the density, of the ray direction and the color (using the color field) [7, 3, 8]. However, such representation alone has a severe limitation: it requires the camera position to be known in order work properly. There have been propositions to estimate the camera pose and to registrate the object deformation. The main difficulty is that the areas without information (with a density value of zero), have uncertain gradient directions. The proposed solutions often require to be close to the line of sight or have a simple scene in order to have a correct reconstruction.

Distance Field Given a 3D position as the input, the density field returns its distance to the nearest neighbour boundary. The most commonly used type of Distance Fields are Signed Distance Fields (SDF) . They are, by construction, able to provide stable bounding surfaces and normal vectors, all while providing residuals and gradient directions. This allows for fast fitting of two shapes by using the Gauss-Newton Method without the need for point matching. Several approaches have been proposed to handle these distance fields with neural fields [6, 10, 2, 11, 1, 5]. However, one of the issues of encoding denisity fields in neural networks is that they are noisy when reconstructing the surfaces by level sets. Some works have proposed to alleviate this issue by allowing the neural fields to present boundary surfaces and by assuming a static density distribution for the signed distance. But this leads to other problems, especially when reconstructing complex shapes with abrupt depth changes. This is due to the fact that these methods are base on calculation of the single surface intersection for each ray and this methods leads to an unstable reconstruction in this context.

2.2 Method

The idea of the authors is to provide a consistent distance field while retaining the expressiveness of the density field by providing gradients where no objects are present. This allows for an improved registration performance from a rough initial camera pose estimation.

Continuing the basics established in section 1.2. It is important to add that $d(\mathbf{p})$ takes the same value for the depth value $d_b(\mathbf{p})$ in the presence of a boundary surface by taking the density $\sigma(\mathbf{p})$ to be 0 outside $(0 < t < d_b)$ and ∞ inside $(d_b \leq t)$. Furthermore, with the simplification of the distance field, we assume that the adopted view direction is continuous, which allows us to recover its value using the gradient of the distance field $\nabla D(\mathbf{p})$:

$$\mathbf{v}_n = \frac{-\nabla D(\mathbf{p})}{\|\nabla D(\mathbf{p})\|_2} \tag{4}$$

$$\nabla D(\mathbf{p}) = \left[\frac{\partial D(\mathbf{p})}{\partial p_x} \frac{\partial D(\mathbf{p})}{\partial p_y} \frac{\partial D(\mathbf{p})}{\partial p_z}\right]. \tag{5}$$

But this assumption doesn't hold in real-case scenarios, as $\mathbf{v}_n$ has discontinuities in practice. In order to find strategies to alleviate this issue, the authors propose to explore the case where the distance field is known in order to derive the density field:

$$\left.\frac{\partial D(\mathbf{r}(t))}{\partial t}\right|_{t=0} = \lim_{\Delta t \to 0} \frac{d(\mathbf{r}(\Delta t), \mathbf{v}) - d(\mathbf{r}(0), \mathbf{v})}{\Delta t} \tag{6}$$

$$= -1 + \left(D(\mathbf{p}) - t_n\right) \sigma\left(\mathbf{p} + t_n \mathbf{v}\right) \tag{7}$$

The authors also propose an expression of $\frac{\partial D(\mathbf{r}(t))}{\partial t}$ using the gradient vector of the distance field $\nabla D(\mathbf{p})$:

$$\frac{\partial D(\mathbf{r}(t))}{\partial t} = \frac{\partial D}{\partial p_x}\frac{\partial p_x}{\partial t} + \frac{\partial D}{\partial p_y}\frac{\partial p_y}{\partial t} + \frac{\partial D}{\partial p_z}\frac{\partial p_z}{\partial t} \tag{8}$$

$$= \nabla D(\mathbf{p}) \cdot \mathbf{v} \tag{9}$$

$$= -\|\nabla D(\mathbf{p})\|_2. \tag{10}$$

Now, using Equations 7 and 10, the density can be recovered with equation 3.

Let the variable t_n that defines the range within which the light transmittance, $T(t_n)$, to 1, effectively setting the lower boundary for the depth value D. In real-case scenarios, t_n can have a value of 0, but this leads to a computational issue when D approaches 0, as σ becomes undefined. To circumvent this problem, the authors propose that t_n is a negligibly small quantity which allow to rewrite σ as follows :

$$D\left(\mathbf{p} - t_n \mathbf{v}\right) \simeq D(\mathbf{p}) + t_n, \tag{11}$$

$$\nabla D\left(\mathbf{p} - t_n \mathbf{v}\right) \simeq \nabla D(\mathbf{p}) \tag{12}$$

$$\sigma(\mathbf{p}) \simeq \frac{1 - \|\nabla D(\mathbf{p})\|_2}{D(\mathbf{p})} \tag{13}$$

Given that the network is differentiable, it can compute the distance $D\left(\mathbf{p}_i\right)$ and gradient vector $\nabla D\left(\mathbf{p}_i\right)$ with independent sampling points $\mathbf{p}_i$ as input to the neural field. However, a distance field is not differentiable at the cusps where the minimum direction switches. The solution proposed by the authors to solve this problem is to add an auxiliary gradient axis w, leading to a hyperspace $[x, y, z, w]$. The gradient components $\frac{\partial D}{\partial w}$ are built in a way that the gradient ∇D satisfies the Equation 13 near the cusps, which removes the wrong densities. The authors also constrain the auxiliary gradient outside of the regions with cusps by adding a penalty term allowing a correct convergence. For the case $\frac{\partial D}{\partial t} > 0$, the proposed heuristic is:

$$\frac{\partial^2 D}{\partial t \partial w} = \alpha \frac{1}{D} \frac{\partial D}{\partial w} \tag{14}$$

For the case $\frac{\partial D}{\partial t} = 0$, they introduce a weight coefficient β defined as follows:

$$\beta = D \left(\frac{\partial D}{\partial t} \right)^2 \frac{\partial D}{\partial w} \tag{15}$$

Finally, the objective function for the shape constraint of the auxiliary gradient w is defined as:

$$L_{\text{const}} = \frac{\lambda_{\text{const}}}{M} \sum_{\mathbf{p} \in \mathcal{P}} \beta \left[\frac{\partial^2 D}{\partial t \partial w} - \frac{\alpha}{D(\mathbf{p})} \frac{\partial D}{\partial w} \right]^2 \tag{16}$$

With $\mathcal{P}$ the sampling points and λ_{const} an hyperparameter of the model.

The final proposition of the authors regards the retroprojection error for the volume rendering. NeRF-based localization methods use the photometric error to render the volume. This error is simply the residual $\|\mathbf{C}(\mathbf{q}) - \hat{\mathbf{C}}(\mathbf{q})\|_2$ from the observed color $\mathbf{C}(\mathbf{q})$, and the estimated color $\hat{\mathbf{C}}(\mathbf{q})$ aggregated by volume rendering for a given pixel $\mathbf{q}$. Such approach can only handle objects with hard surfaces as well as local regions with smooth color changes. The NeDDF on the other hand holds the information, for each sampling point, of the object's direction as well as it's approximate distance from said point. This allows the computation of the pseudo-correspondence point and estimate the camera pose using the reprojection error.

In order to calculate the correspondance points using the color information, the proposition is to penalize the color change in the distance gradient direction for blanck regions. For a point $\mathbf{p}_i$, camera depth t_i and a viewing direction $\mathbf{v}$, let the output be color $\mathbf{c}_i$, distance D_i, and distance gradient ∇D_i. The penalty takes $L_{\text{blank}} = \sum_i \left\| \nabla \mathbf{c}_i \left(\nabla D_i \right)^T \right\|_2$. Since $\|\nabla D_i\|_2$ takes small values inside the object, this penalty is applied outside outside the object. Training the network with this penalty makes it possible to obtain the nearest neighbor object's color, direction, and distance from sampling points outside the object.

To compute the pseudo-correspondence points for each ray we use the points closer to the observed color than the estiamted one and we combine nearsurface points by focusing weights closer to the color and the distance between them. In order to do this, the weight g_i of the sampling point $\mathbf{p}_i$ is defined as follows:

$$g_i = \text{softmax} \left(-\lambda_D \frac{D_i \|\nabla D_i \times \mathbf{v}\|_2}{t_i} - \lambda_c \|\mathbf{C}(\mathbf{q}) - \mathbf{c}_i\|_2 \right) \tag{17}$$

With λ_D and λ_c hyperparameters of the model. Finaly, the authors present the reprojection error that measures the distance $\|\mathbf{q} - \hat{\mathbf{q}}\|_2$ between the pixel coordinates of the ray $\mathbf{q}$ and the projected pseudo-correspondence point $\hat{\mathbf{q}}$.

This work allows the NeDDF to have both the propories of density and distance fields while retaining the reconstruction quality of standard NeRF but also enabling for more stable meshes as well as being more robust to various camera poses.

References

[1] Matan Atzmon and Yaron Lipman. “Sal: Sign agnostic learning of shapes from raw data”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2020, pp. 2565–2574.

[2] Julian Chibane, Aymen Mir, and Gerard Pons-Moll. “Neural Unsigned Distance Fields for Implicit Function Learning”. In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2020, pp. 21638–21652.

[3] Zhengqi Li et al. “Neural Scene Flow Fields for Space-Time View Synthesis of Dynamic Scenes”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2021, pp. 6498–6508.

[4] Ben Mildenhall et al. “Nerf: Representing scenes as neural radiance fields for view synthesis”. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 2020, pp. 405–421.

[5] Michael Oechsle, Songyou Peng, and Andreas Geiger. “UNISURF: Unifying Neural Implicit Surfaces and Radiance Fields for Multi-View Reconstruction”. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. 2021, pp. 5589–5599.

[6] Jeong Joon Park et al. “Deepsdf: Learning continuous signed distance functions for shape representation”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2019, pp. 165–174.

[7] Keunhong Park et al. “Nerfies: Deformable Neural Radiance Fields”. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (2021), pp. 5865–5874.

[8] Albert Pumarola et al. “D-nerf: Neural radiance fields for dynamic scenes”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2021, pp. 10318–10327.

[9] Itsuki Ueda et al. *Neural Density-Distance Fields*. 2022. arXiv: `2207.14455 [cs.CV]`.

[10] Peng Wang et al. “NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-view Reconstruction”. In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2021.

[11] Lior Yariv et al. “Multiview Neural Surface Reconstruction by Disentangling Geometry and Appearance”. In: *Advances in Neural Information Processing Systems (NeurIPS)*. Vol. 33. 2020.