# 3D Shape Modeling with Adaptive Centroidal Voronoi Tesselation on Signed Distance Field

Diego Thomas, Jean-Sébastien Franco, Edmond Boyer

# 3D Shape Modeling with Adaptive Centroidal Voronoi Tesselation on Signed Distance Field

Diego THOMAS[1,a)] Jean-Sebastien FRANCO[2,b)] Edmond BOYER[2,c)]

## Abstract

Volumetric shape representations have become ubiquitous in multi-view reconstruction tasks. They often build on regular voxel grids as discrete representations of 3D shape functions, such as SDF or density. Despite their proven efficiency, voxel representations come with the precision versus complexity tradeoff. To alleviate this problem hierarchical octrees, hash maps or static tetrahedral structure have been proposed recently. In this paper we investigate an alternative discretization strategy with an adaptive Centroidal Voronoi Tesselation (CVT). CVTs allow to better partition the observation space with respect to shape occupancy and to focus the discretization around shape surfaces. To leverage this discretization strategy for multi-view reconstruction, we introduce an approximate implicit CVT optimization that does not require explicit cells and adapts to the reconstructed SDF field. We adapt and reformulate the 3D shape reconstruction framework for our proposed CVT structure and achieve state-of-the-art reconstruction results on public datasets.

## 1. Introduction

The 3D digitization of real-world objects is crucial for future technologies and has motivated extensive research. Among the primary solutions, multi-view capture systems have arisen as key tools to generate high-quality shape and appearance models of 3D scenes. However, the reconstruction of detailed geometry from multiple high-resolution images remains a challenging task due to the inherent ambiguities and complexity in the visual observations.

Volumetric shape representations are increasingly prevalent in multi-view 3D reconstruction approaches, e.g. [10], [14], [21], [23], [24]. This is in part due to their ability to relate shape properties to image observations through differential rendering [12]. This has been extensively leveraged, in particular with networks, typically MLPs [8], which are trained to model shape geometry or appearance that best explain the image observations under photometric losses.

In the seminal work NerF [14] and the follow-up [2], [3], [4], [26], volumetric radiance fields are estimated by integrating color and opacity along pixel rays. The problem is difficulty to capture higher detail surface extraction. Subsequent works [15], [16], [21], [24], address this by explicitly parameterizing the surface and its properties with signed distance fields (SDF). Hybrid methods that combine explicit SDF grids with shallow networks for color offer benefits in both geometry and appearance modelling [5], [6], [7], [17], [19], [23]. Yet, most of such methods rely on some form of regular axis-aligned grids to discretize 3D observation spaces and are therefore sensitive to the inherently poor quality-to-parsimony trade-off of these representations. Moreover, regular grids result in sub-optimal meshes when paired with the Marching Cubes algorithm [13] ubiquitously used for explicit mesh surface conversion.

Voxel grids uniformly discretize the observation space, regardless of the shape's location. Consequently, increasing resolution specifically near the shape surface requires non-trivial specializations. Octrees offer hierarchical space discretization [10], [11], [22], [25], but their dynamic update during optimization and raymarching is cumbersome. Tetrahedral meshes built on dense point clouds obtained with

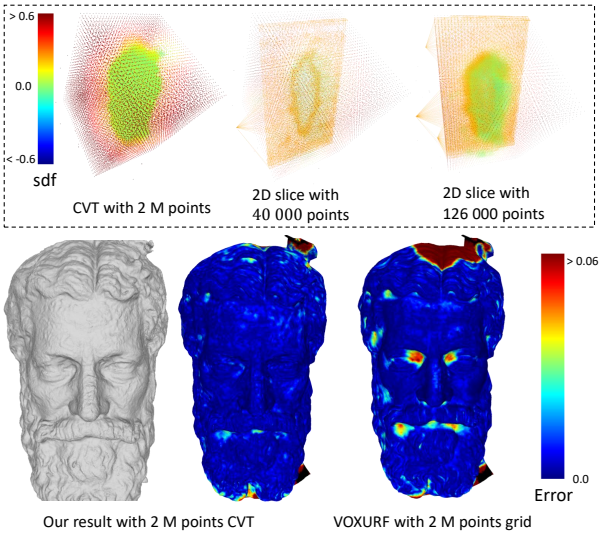


Fig. 1: Our proposed method reconstruct detailed 3D surfaces with using Centroidal Voronoi Tessellation that adapts to the reconstructed geometry. With equivalent discretisation level, our method achieves an order of magnitude higher level detail in the 3D reconstructed mesh.

[1] Kyushu University, Japan
[2] INRIA, France
[a)] thomas@ait.kyushu-u.ac.jp
[b)] jean-sebastien.franco@inria.fr
[c)] edmond.boyer@inria.fr

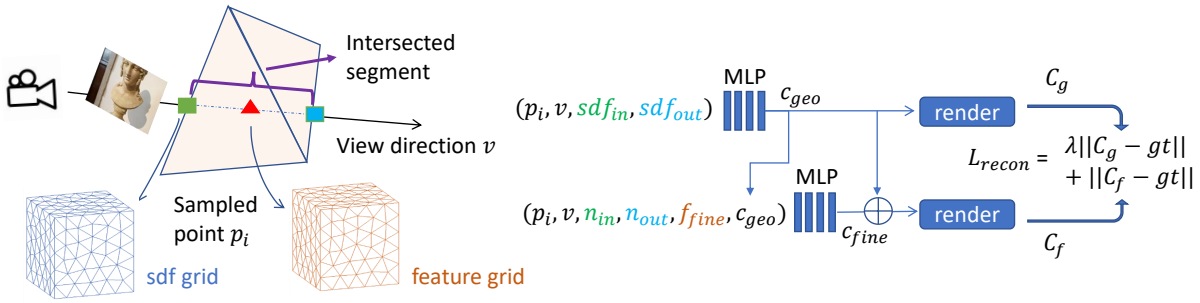


Fig. 2: We propose a volumetric optimization framework that combines explicit SDF fields with shallow color networks, in order to estimate 3D shape properties over tetrahedral grids.

structure-from-motion algorithms [18] allow to focus discretization around the surface[9]. However, they focus on volumetric radiance fields reconstruction and cannot reconstruct the 3D surface, dynamic update of the discretization is not possible and computational time is heavy. Instead, we explore a hierarchical tetrahedral discretization guided by the Centroidal Voronoi Tesselation (CVT) algorithm. CVTs yield provably optimal discretizations and exhibit noteworthy advantages in our context: (i) Efficient ray marching along rays through tetrahedra; (ii) The ability to hierarchically up-sample and deform tetrahedral grids in adaptive fashion w.r.t the encoded shape surface.

As conventional 3D convolution or automatic differentiation do not extend to such non-uniform cell complexes, a develop a complete representation and methodology for optimization. Our approach jointly optimizes a hierarchical CVT discretization with an associated neural field. Two shallow color networks are trained to predict the colors at sample points based on the interpolated learnable features. We show that feature extraction and gradient backpropagation along rays can be efficiently performed over the tetrahedral dual of the CVT. Our hierarchical approach upsamples the CVT, after neural field convergence, at increasing levels of details, with a tenfold difference in grid resolution between subsequent levels. Our experiments demonstrate that this strategy yields significantly more reconstruction detail when compared to SOTA techniques with equivalent time and primitive budgets.

Our main contributions are: (1) introducing CVT discretization for neural fields in multi-view reconstruction; (2) qn implicit CVT optimization method that adapts to the optimized SDF field; (3) proposing an associated fast optimization framework with.

## 2. Method

Given a set of multiple images, our method jointly optimizes a SDF field and a feature field, discretized on a hierarchical CVT, and two view-dependent shallow color networks to predict view-dependent color at any 3D location (Figure 2). SDF and feature values are optimized by backpropagating a photometric loss from the sampled points to the CVT sites. After convergence, the CVT is iteratively refined and up-sampled non-uniformly with respect to the current estimate of the SDF field, increasing therefore the shape resolution at the vicinity of the shape surface.

### 2.1 Coarse-to-fine Centroidal Voronoi Tesselation

Central to our proposed method is densifying the continuous SDF discretization near the surface. We begin with a coarse initial CVT[*1], to which we incorporate the center of each camera as additional sites to in order to speed-up ray-marching operations.

After the optimization of the color network parameters and of the SDF and feature values within the CVT, we upsample the discretization by adding a point at the center of each surface-crossing edge of the dual of the CVT. The CVT energy is then minimized, at each up-sampling iteration [20][*2]. This process is repeated until the expected level of details is reached.

#### 2.1.1 SDF-aware implicit CVT optimization

Recomputing the topology of the CVT (*i.e.* the cells boundaries) at each step of the CVT optimization is time consuming. We propose to build an approximate CVT implicitly without explicitly computing the topology. We also adapt the optimization to a given SDF field.

Each Voronoi cell is a convex polygon. Then given a site position $s_i$ and a ray $\mathbf{r}$, the distance $d(\mathbf{s_i}, \mathbf{r})$ from the site to the border of the Voronoi cell in the ray direction is obtained as the minimum distance between the site and the intersection points between the ray and each bisector plane of the adjacent sites.

Our key observation is that in an optimized CVT, for each ray that starts from the site the distance from the site to the border of the Voronoi cell is the same in both directions. We define our CVT loss by randomly rotating the cartesian basis with two random angles $(\Theta, \Phi)$. We write the rotated basis as $(\mathbf{e_0})(\Theta, \Phi), \mathbf{e_1})(\Theta, \Phi), \mathbf{e_2})(\Theta, \Phi)$. Then the CVT loss is

$$L_{CVT} = \frac{1}{2} \sum_{s_i} \sum_{j=0,1,2} (d(\mathbf{s_i}, \mathbf{e_j}(\Theta, \Phi)) - d(\mathbf{s_i}, -\mathbf{e_j}(\Theta, \Phi)))^2.$$

We also add an SDF regularisation that moves adjacent sites that cross the surface such that the bisector planes lie

[*1] Generated from a $16 \times 16 \times 16$ grid in our experiments.
[*2] This energy makes the sites coincide with the centroids of the associated cells.

on the surface. Our proposed SDF loss for the CVT is

$$L_{SDF} = \frac{1}{2}\sum_{s_i}\sum_{\mathbf{s_j}\in N(\mathbf{s_i})}\sigma(\mathbf{s_j},\mathbf{s_j}), \quad (1)$$

where $\sigma(\mathbf{s_j},\mathbf{s_j}) = 0$ if $\mathbf{s_j}$ and $\mathbf{s_j}$ are not adjacent or the SDF signs are the same, $\sigma(\mathbf{s_j},\mathbf{s_j}) = (d(\mathbf{s_i},\mathbf{s_j}-\mathbf{s_i}) - d(\mathbf{s_j},\mathbf{s_i}-\mathbf{s_j})^2$ otherwise. We implement the losses on the GPU, compute the explicit gradients and use PyTorch Adam optimizer to minimise $L_{CVT} + L_{SDF}$ with 300 iterations. After the CVT is optimized, the Delaunay tetrahedra are computed to restart the SDF and features optimization.

### 2.2 Differentiable rendering in tetrahedral mesh

Differentiable rendering is a key component of volumetric reconstruction strategies. In such rendering, efficient sampling of points along pixel viewing lines is crucial. However, doing efficient sampling on non-uniform tetrahedral grids require specific algorithms.

We adapt the 32 bits tetrahedral structure proposed in [1] for fast ray-marching. The key idea is to encode the index of the 4th vertex of each tetrahedron with the XOR operator. Then given the indices of 3 of the 4 vertices of the tetrahedron the remaining index is obtain with XOR operation. We walk through the tetrahedral mesh and along a viewing line by: (i) finding the exit face in the current tetrahedron; (ii) identifying the next tetrahedron and the new vertices in the next tetrahedron with the XOR operator. To find the exit face we project all vertices in the plane centered at the camera center and which normal equals the ray direction vector. Then the exit face is the face that contains the origin in the projected coordinate system (the entry face is not counted).

#### 2.2.1 Volumetric rendering

The 3D ray $(\mathbf{o},\mathbf{v})$ originates from the camera center $\mathbf{o}$ in the direction $\mathbf{v}$. When intersecting the tetrahedral grid, it is split into $n$ segments $\{s(t) = [\mathbf{in}(t) : \mathbf{out}(t)] | 0 \le t \le n\}$ with non-null contribution to the accumulated color using the output of our proposed ray-marching algorithm. For each segment we compute transmittance, accumulated transmittance and coarse and refined colors at the middle point. We then follow the method proposed in [21] to compute coarse and refined colors $\mathbf{C_i^{geo}}$ and $\mathbf{C_i^{fine}}$ for the ray. Given the ground truth color $\mathbf{C_i}$ at the ray, we get the following photometric data term for the SDF optimization:

$$E_{color} = \sum_{i\in[1:N]} w_{geo}H(\mathbf{C_i} - \mathbf{C_i^{geo}}) + H(\mathbf{C_i} - \mathbf{C_i^{fine}}),$$

where $H$ is the Huber loss and $w_{geo}$ is a weight (1 at the coarser stage then 0.5).

### 2.3 SDF field regularization

While fully implicit surface representations are naturally regularized by the weights of the neural network, special attention is required to regularize discrete SDF fields.

#### 2.3.1 Eikonal loss

The Eikonal loss is used to regularize the discretized SDF field and makes it resemble a true SDF field with which the norm of the gradient $\nabla sdf$ would be 1.

$$L_{ek} = \sum_t (\|\nabla \mathbf{sdf_t}\|_2 - 1)^2, \quad (2)$$

where the index $t$ represents the index of a sampled point.

Since SDF values are interpolated within a tetrahedron, we can express the gradient of the SDF function within a tetrahedron as a function of the values at the tetrahedron's vertices and of the spatial gradients $\nabla \mathbf{w_i}$ of the interpolation weights $w_i$. By solving a linear system in each tetrahedron we compute the gradient vector associated to each tetrahedron. These gradients vectors are averaged to the summits to compute the eikonal loss.

#### 2.3.2 Normal Smoothing

As proposed in [23] we use a smoothing regulator that minimize the distance between the gradients of the sdf and the gradients of the smoothed sdf values.

$$L_{reg} = \sum_t \|\nabla \mathbf{sdf_t} - \nabla \mathbf{sdf_t}^{smooth}\|. \quad (3)$$

We compute the smoothed SDF values on the CVT by using weighted average of SDF values of K-nearest sites in the current CVT and the K-nearest sites in the CVT at previous level[*3]. We also use the total variation loss defined on the edges of the tetrahedral mesh, with Gaussian weights of the length of the edges. The final SDF gradient writes:

$$\frac{\partial sdf}{\partial t} = \frac{\partial E_{color}}{\partial t} + w_{ek}\frac{\partial L_{ek}}{\partial t} + w_{reg}\frac{\partial L_{reg}}{\partial t} + w_{tv}\frac{\partial L_{TV}}{\partial t},$$

where $w_{ek}$, $w_{reg}$ and $w_{tv}$ are weight factors.

## 3. Experiments

We evaluate the ability of our method to model detailed 3D shapes compared with the state-of-the-art methods NeuS [21], Neus2 [22] and Voxurf [23]. We use the code provided by the authors for all methods and run our experiments on an RTX3090 GPU. We qualitatively and quantitatively evaluate our method on a subset of the BlendedMVS dataset.

We used 7 scenes of the blendedMVS dataset. Two different scenarios occur: small scenes where objects occupy most of the bounding box, which favors in principle uniform discretizations. Second, larger scenes with significant amount of empty space in the bounding box (stone or durian) to confirm the advantage of our adaptive discretization.

For fair evaluation we also compare our method with VOXURF at equivalent levels of discretization, i.e., with similar amounts of points in the grid. Our method usually terminates with about 2M points in the CVT, which corresponds to running VOXURF with a uniform grid of $128 \times 128 \times 128$ voxels. We reconfigure the authors implementation to this goal, all else being equal. We also compare with the $256^3$ version.

Table 1 shows the quantitative evaluation compared to

[*3] The K-nearest neighbors are computed only once at each upsampling step using the corresponding KD-trees.

Table 1: Average geometric accuracy *Acc* (cm) (lower is better) and intersection of union (IoU) (higher is better) obtained with our method, NeuS 2 and Voxurf, for each of the 7 test scenes. We highlight the **best** and second values.

| | NeuS | | NeuS2 | | Voxurf 2$M$ | | Voxurf 16$M$ | | Ours 2$M$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Acc ↓ | IoU ↑ | Acc ↓ | IoU ↑ | Acc ↓ | IoU ↑ | Acc ↓ | IoU ↑ | Acc ↓ | IoU ↑ |
| Dog | 1.50 | 0.78 | 1.70 | **0.95** | 1.65 | 0.92 | **0.95** | 0.86 | 1.94 | 0.92 |
| Bear | 2.08 | **0.99** | **1.94** | 0.83 | 3.47 | 0.78 | 2.60 | 0.79 | 3.00 | 0.87 |
| Clock | 2.44 | 0.86 | 1.73 | 0.97 | 1.54 | **0.99** | **0.96** | **0.99** | 1.48 | **0.99** |
| Durian | 262.6 | **0.53** | 15.6 | 0.39 | 369.1 | 0.28 | 272.6 | 0.38 | **14.1** | 0.48 |
| Man | 2.25 | 0.62 | **2.03** | 0.52 | 4.21 | 0.57 | 2.58 | 0.63 | **2.03** | **0.73** |
| Sculpture | 1.77 | 0.90 | 1.28 | 0.95 | 1.36 | 0.94 | **1.15** | 0.95 | 1.71 | **0.96** |
| Stone | 10.83 | 0.64 | **3.86** | 0.25 | 41.3 | 0.49 | 31.4 | 0.50 | 8.21 | **0.71** |
| Avg | 40.44 | 0.76 | **4.05** | 0.69 | 60.38 | 0.71 | 44.64 | 0.72 | 4.64 | **0.79** |

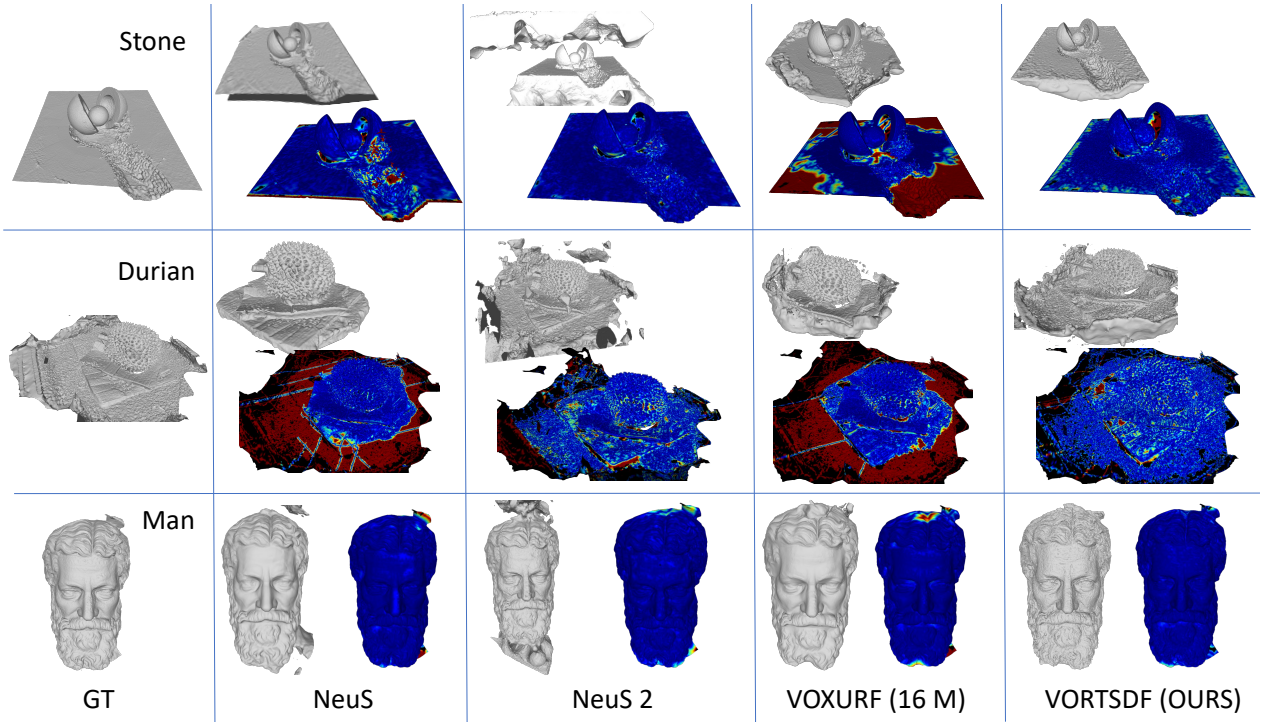


Fig. 3: Comparative results we obtained with our method, NeuS and Voxurf on data "Stone", "Durian" and "Man" of BlendedMVS. We output the final 3D meshes using Marching Cubes (MC) for NeuS and Voxurf and our method. We also show errors from ground truth meshes to predicted meshes as heatmaps.

NeUS and Voxurf. Our method almost always improves the accuracy of the reconstructed 3D mesh compared to these SOTA methods. As expected, we observe more significant gains with large scenes for which uniform space discretizations are not well suited. This confirms that using an adaptive CVT to support the SDF field optimization is an effective solution that yields higher frequency details in the reconstructed geometry and appearance.

Our experimental results also show that, at similar levels of discretization, our method retrieves reconstructed meshes of significantly better quality. Our method even obtains better results than VOXURF when using an order of magnitude less discretization points. Our method is also able to produce significantly fewer artifacts than other methods as shown by the IoU results.

Figure 3 shows qualitative comparisons. They demonstrate that denser discretizations around the surface can effectively yield higher frequency details and less outliers than other methods.

## 4. Conclusion

We propose a novel method to reconstruct 3D geometry of a target scene from a set of multi-view images by optimizing a SDF field on a Centroidal Voronoi Tesselation. We formulate the optimization framework over the CVT and its dual tetrahedral mesh, designing an efficient framework to output detailed 3D shapes with competitive computation times. Our experimental results validate the key ideas in our proposed method and demonstrate that at equivalent discretization level we can achieve a significantly higher level of extracted detail in a majority of situations, compared to competitive approaches.

## References


[1] Aman, A., Demirci, S. and Güdükbay, U.: Compact tetrahedralization-based acceleration structures for ray tracing, *Journal of Visualization*, Vol. 25, No. 5, pp. 1103–1115 (2022).

[2] Barron, J. T., Mildenhall, B., Tancik, M., Hedman, P., Martin-Brualla, R. and Srinivasan, P. P.: Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields, *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 5855–5864 (2021).

[3] Barron, J. T., Mildenhall, B., Verbin, D., Srinivasan, P. P. and Hedman, P.: Mip-nerf 360: Unbounded anti-aliased neural radiance fields, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5470–5479 (2022).

[4] Barron, J. T., Mildenhall, B., Verbin, D., Srinivasan, P. P. and Hedman, P.: Zip-NeRF: Anti-Aliased Grid-Based Neural Radiance Fields, *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (2023).

[5] Chen, A., Xu, Z., Geiger, A., Yu, J. and Su, H.: Tensorf: Tensorial radiance fields, *European Conference on Computer Vision (ECCV)*, Springer, pp. 333–350 (2022).

[6] Chen, A., Xu, Z., Wei, X., Tang, S., Su, H. and Geiger, A.: Factor fields: A unified framework for neural fields and beyond, *arXiv preprint arXiv:2302.01226* (2023).

[7] Fridovich-Keil, S., Meanti, G., Warburg, F. R., Recht, B. and Kanazawa, A.: K-planes: Explicit radiance fields in space, time, and appearance, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12479–12488 (2023).

[8] Kellnhofer, P., Jebe, L. C., Jones, A., Spicer, R., Pulli, K. and Wetzstein, G.: Neural lumigraph rendering, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4287–4297 (2021).

[9] Kulhanek, J. and Sattler, T.: Tetra-NeRF: Representing Neural Radiance Fields Using Tetrahedra, *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (2023).

[10] Li, Z., Müller, T., Evans, A., Taylor, R. H., Unberath, M., Liu, M.-Y. and Lin, C.-H.: Neuralangelo: High-Fidelity Neural Surface Reconstruction, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2023).

[11] Liu, L., Gu, J., Lin, K. Z., Chua, T.-S. and Theobalt, C.: Neural Sparse Voxel Fields, *NeurIPS* (2020).

[12] Lombardi, S., Simon, T., Saragih, J., Schwartz, G., Lehrmann, A. and Sheikh, Y.: Neural Volumes: Learning Dynamic Renderable Volumes from Images, *ACM Trans. Graph.*, Vol. 38, No. 4, pp. 65:1–65:14 (2019).

[13] Lorensen, W. E. and Cline, H. E.: Marching cubes: A high resolution 3D surface construction algorithm, *Seminal graphics: pioneering efforts that shaped the field*, pp. 347–353 (1998).

[14] Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R. and Ng, R.: NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis, *European Conference on Computer Vision (ECCV)* (2020).

[15] Oechsle, M., Peng, S. and Geiger, A.: Unisurf: Unifying neural implicit surfaces and radiance fields for multi-view reconstruction, *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 5589–5599 (2021).

[16] Park, J. J., Florence, P., Straub, J., Newcombe, R. and Lovegrove, S.: Deepsdf: Learning continuous signed distance functions for shape representation, *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp. 165–174 (2019).

[17] Reiser, C., Peng, S., Liao, Y. and Geiger, A.: Kilonerf: Speeding up neural radiance fields with thousands of tiny mlps, *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 14335–14345 (2021).

[18] Schönberger, J. L. and Frahm, J.-M.: Structure-from-Motion Revisited, *Conference on Computer Vision and Pattern Recognition (CVPR)* (2016).

[19] Sun, C., Sun, M. and Chen, H.-T.: Direct voxel grid optimization: Super-fast convergence for radiance fields reconstruction, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5459–5469 (2022).

[20] Wang, L., Hétroy-Wheeler, F. and Boyer, E.: A Hierarchical Approach for Regular Centroidal Voronoi Tessellations, *Computer Graphics Forum*, Vol. 35, No. 1, pp. 152–165 (2016).

[21] Wang, P., Liu, L., Liu, Y., Theobalt, C., Komura, T. and Wang, W.: NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-view Reconstruction, *arXiv preprint arXiv:2106.10689* (2021).

[22] Wang, Y., Han, Q., Habermann, M., Daniilidis, K., Theobalt, C. and Liu, L.: NeuS2: Fast Learning of Neural Implicit Surfaces for Multi-view Reconstruction, *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (2023).

[23] Wu, T., Wang, J., Pan, X., Xu, X., Theobalt, C., Liu, Z. and Lin, D.: Voxurf: Voxel-based Efficient and Accurate Neural Surface Reconstruction, *International Conference on Learning Representations (ICLR)* (2023).

[24] Yariv, L., Kasten, Y., Moran, D., Galun, M., Atzmon, M., Ronen, B. and Lipman, Y.: Multiview Neural Surface Reconstruction by Disentangling Geometry and Appearance, *Advances in Neural Information Processing Systems*, Vol. 33 (2020).

[25] Yu, A., Li, R., Tancik, M., Li, H., Ng, R. and Kanazawa, A.: Plenoctrees for real-time rendering of neural radiance fields, *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 5752–5761 (2021).

[26] Zhang, K., Riegler, G., Snavely, N. and Koltun, V.: NeRF++: Analyzing and Improving Neural Radiance Fields, *arXiv:2010.07492* (2020).