Signal, Image and Video Processing (2025) 19:694
https://doi.org/10.1007/s11760-025-04304-3

ORIGINAL PAPER

SPSN-MVPS: Shifting points, silhouettes and neural inverse for solving multiview photometric stereo

Lyes Abada[1] · Tarek Gacem[1] · Aimen Said Mezabiat[1] · Saad allah Bourzam[1] · Omar Chouaab Malki[1] · Mohamed Mekkaoui[1]

Received: 27 February 2025 / Revised: 25 April 2025 / Accepted: 25 May 2025

Abstract

Three-dimensional object reconstruction from images is a core area in computer vision, primarily aimed at deriving the geometric structure of an object from images. A prominent technique within this domain is multi-view photometric stereo, which leverages multiple images captured from different camera viewpoints under varying illumination conditions to accurately infer surface geometry. This paper proposes a novel approach to 3D reconstruction by utilizing the displacement of spatial points to train an implicit network. The loss function is principally based on surface gradients and the Shape from Silhouettes, aiming to address the challenge of 3D reconstruction by multi-view photometric stereo using a neural inverse network. The main concept of this method is to train a multilayer neural network to determine the surface of an object. This technique integrates surface orientation data and surfaces derived from image silhouettes to form a robust loss function. It leverages the movement of points along direction vectors, utilizing surface gradients and shape from silhouette domain, focusing exclusively on the essential information required to train the neural network. A detailed experimental section will be presented in this paper to showcase the obtained results, highlighting the high performance of our new approach. We demonstrate that the proposed method outperforms existing techniques in terms of quality and processing time across several tests. Our code will be available on github after acceptance at : https://github.com/lyabada/SPSN/.

Keywords 3D Reconstruction · Photometric Stereo · Multi-View · Surface Gradient · Shape from Silhouette · Shifted Points

1 Introduction

3D reconstruction is a key area of computer vision focused on generating three-dimensional models of objects from images. These techniques have widespread applications in fields such as medicine [1], Environmental protection [2], robotics [3], object detection [4], and several other sectors. Over time, various approaches have been developed to tackle the challenges of 3D reconstruction, including deep learning [5] and meta-heuristic techniques [6].

One prominent technique for 3D reconstruction is photometric stereo [7], which generates a 3D model from multiple images captured from a fixed camera position. However, this approach typically results in the reconstruction of only one side of the object due to a lack of images from other viewpoints. To overcome this limitation, photometric stereo has evolved into a more general method known as multi-view photometric stereo (MVPS) [8–10]. This method allows for the reconstruction of an object from images captured from multiple angles, leading to a more comprehensive and accurate model.

Multi-view 3D reconstruction has advanced rapidly, improving the modeling of complex objects and scenes. Methods fall into geometry-based, data-driven, and differentiable rendering approaches, benefiting from better image acquisition, increased computing power, and AI advances. This paper introduces a novel approach that prioritizes spatial points over image pixels, leveraging an initial Shape from Silhouette (SfS) object and implicit neural representations to improve surface continuity and quality.

The main contribution of this paper :

✉ Lyes Abada
labada@usthb.dz

[1] Artificial Intelligence Laboratory (LRIA), Computer Science Faculty, University of Science and Technology Houari Boumediene (USTHB), BP 32, Bab Ezzouar 16111, Algiers, Algeria

Published online: 05 June 2025

- A novel 3D reconstruction approach using an inverse network that starts from surface points instead of image pixels.
- Relies solely on silhouettes and surface normals, reducing computation and improving applicability to low-detail objects.
- Enhances processing by integrating Shape-from-Silhouette (SFS) into the objective function, ensuring faster convergence.

2 Related works

In the literature, various techniques for 3D reconstruction using multi-view can be found, including Neural Radiance Fields (NeRF) [8], which model a scene radiance field by estimating color and density from 3D coordinates and viewing directions. NeRF produces highly realistic renderings, accurately capturing complex environments from multiple viewpoints. Among NeRF extensions, UNISURF [9] enhances efficiency by leveraging sparse representations, reducing computational overhead while maintaining quality. It selectively samples relevant scene regions and integrates adaptive sampling and multi-scale prediction to further lower complexity. Combining photometric stereo with multi-view reconstruction improves accuracy by using light intensity variations to infer surfaces. Photometric stereo captures images under different lighting conditions, refining shape and surface details [7]. PSNeRF [12] integrates NeRF, UNISURF, and photometric stereo, using neural networks to model radiance fields while enhancing computational efficiency. By incorporating photometric stereo, PSNeRF achieves more detailed and precise 3D reconstructions. PSNeRF improves 3D reconstruction by integrating photometric stereo, enhancing fine detail and texture capture beyond traditional NeRF. However, its reliance on color and surface normals increases computational cost and slows convergence [12]. Shape from Silhouette (SFS) is another key technique, reconstructing 3D models by analyzing object contours from multiple viewpoints. It defines shape by identifying the common volume enclosed by these silhouettes [13].

3 Three-dimensional reconstruction based on Shape from Silhouette (SFS)

Silhouettes define the region occupied by an object in an image, typically represented as binary masks distinguishing the object from the background. In Shape from Silhouette (SFS), these masks are used to reconstruct 3D shapes from multiple viewpoints [13] or enhance other reconstruction techniques [10]. The core principle involves projecting 2D silhouettes into 3D space to define object volume, relying on the visual hull [13], which represents the intersection of silhouette-derived volumes. Silhouette-based methods are effective for opaque object reconstruction, offering simplicity and real-time capability. However, they struggle with fine surface details and concave regions due to their reliance on contour information. The proposed method builds on Shape from Silhouette (SFS) techniques to address some limitations of traditional multi-view photometric stereo methods. The following section provides a detailed description of our approach.

4 Gradient and Silhouette for neural multi view photometric stereo

In this section, we present a detailed overview of the proposed method, which exploits surface orientation derived from gradients or normal vectors, combined with object surface information obtained through Shape from Silhouette. The core objective of this approach is to train a Multi-Layer Perceptron (MLP) neural network to classify points as belonging to the surface. To achieve this, several strategies have been explored. A key challenge in these techniques lies in determining the most effective loss function to optimize network training. We do not focus on the specific architecture of the neural network, as various effective architectures have already been proposed. Instead, our primary emphasis is on the loss function, which forms the foundation of our improvements.

4.1 The neural network architecture for object occupancy

Unlike methods that employ a cascaded training of two networks; one for geometry and another for appearance, such as PSNeRF [12] and UNISURF [9], our approach utilizes a single MLP dedicated exclusively to the geometric aspect. This model predicts the occupancy of each point in space without considering appearance (e.g., color), which is known to present several drawbacks. For instance, the same point on an object can exhibit different colors and brightness levels when observed from different viewpoints [9, 12].

The MLP architecture employed in our approach is identical to the geometric component proposed in previous studies [9, 12], demonstrating its effectiveness. A detailed description of this architecture will be provided in the experimentation section.

Our model takes a point in space as input and returns its occupancy value: 1 if inside the object, 0 if outside, and 0.5 if on the surface, as shown in Figure 1.

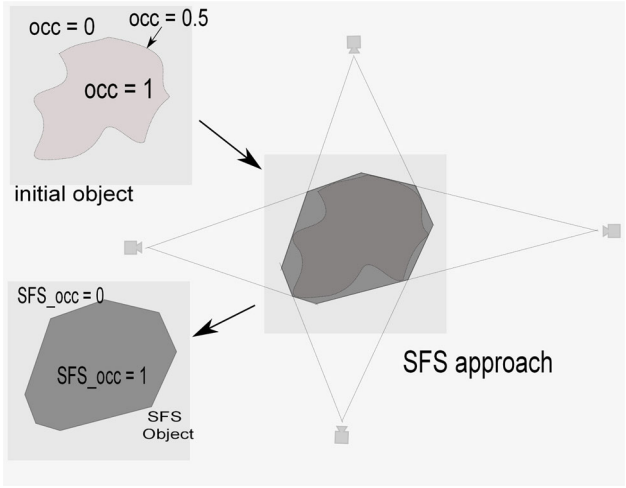


Fig. 1 SFS occupancy (density)

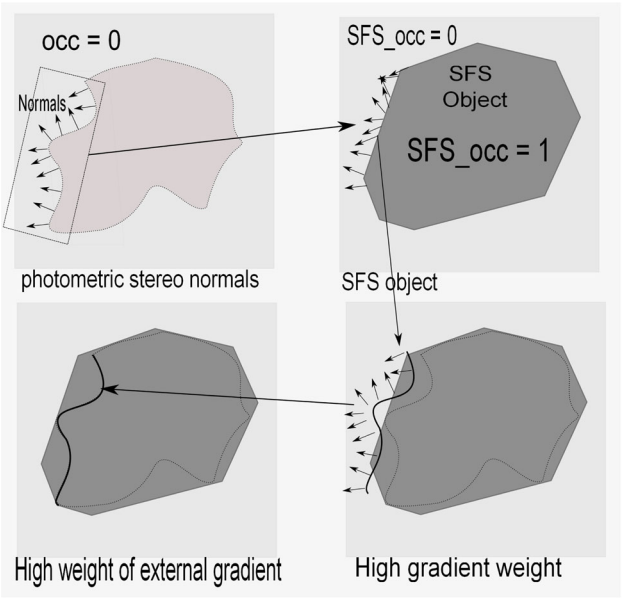


Fig. 2 Surface enhancement using the gradient (normals vectors)

4.2 Loss function

The loss function plays an important role in this category of method. Our approach integrates the normal field from single-view photometric stereo to identify a surface whose normal vectors align with those computed using photometric stereo. This process requires an initial surface to derive the normals.

In most techniques, training begins with either an empty object or a sphere, but both are far from the final object. We propose to constrain the model to generate an initial object using the occupancy provided by SFS, which is then refined by the surface gradient. Our loss function consists of two components: the first is associated with the object, computed using the Shape from Silhouette (SFS) technique, and the second adjusts the surface orientation based on the gradient. This second component is similar to the integration of the normal field in single-view reconstruction, thereby improving the surface orientation (see Figure 2).

4.2.1 SFS loss function

Shape from Silhouette (SFS) generates a 3D object by intersecting the silhouettes from each view in space. The object produced by SFS encompasses the ground truth object, meaning that all occupied points of the ground truth object are located precisely within the volume of the object generated by SFS. Figure 1 illustrates an explanatory diagram of 3D reconstruction using Shape from Silhouette (SFS). A point in space is considered occupied if and only if it belongs to all silhouettes of the object. The first step involves training the model on the object generated by SFS. We know that the object is precisely located within the occupied region, and no occupied points exist outside of it. However, there are non-occupied points that are incorrectly classified as occupied by SFS. To address this, it is essential to give more weight to the outer region (occ=0) compared to the inner region (occ=1). We can summarize the SFS loss function by Equation 1.

$$\mathcal{L}_{sfs} = \mathcal{W}_{in}\mathcal{L}_{in} + \mathcal{W}_{out}\mathcal{L}_{out} \quad \setminus \quad \mathcal{W}_{in} << \mathcal{W}_{out} \tag{1}$$

In this equation, $\mathcal{L}_{out}$ represents the loss function for the points located outside the object P_{out} generated by SFS. $\mathcal{L}_{in}$ denotes the loss function for the points within the area P_{in} considered to be occupied by SFS. This value holds limited importance (low weight $\mathcal{W}_{in}$) because its accuracy is not consistent across all points.

But it is necessary because it drives the surface toward the boundaries of the unoccupied area. A very low value causes the model to generate an empty surface, while a very high value results in the occupied surface extending beyond the area considered unoccupied by the SFS. Therefore, it is essential to assign the lowest value that guides the surface toward the unoccupied area without exceeding it.

$\mathcal{L}_{in}$ and $\mathcal{L}_{out}$ are determined by Equations 2 and 3, where occ_{sfs} represents the occupancy of the SFS and occ_{pred} represents the occupancy provided by the network.

$$\mathcal{L}_{in} = \sum_{p \in P_{int}} \|occ_{sfs} - occ_{pred}\|_2 \tag{2}$$

$$\mathcal{L}_{out} = \sum_{p \in P_{ext}} \|occ_{sfs} - occ_{pred}\|_2 \tag{3}$$

4.2.2 Gradient loss function

The NeRF technique primarily uses color to determine the position of points in space relative to different views. The authors of this approach clearly demonstrate that color is highly sensitive to lighting and sensor conditions, yet it remains essential for multi-view reconstruction. The $NeRF^{+N}$, $UNISURF^{+N}$ and others similar techniques introduce the concept of photometric stereo (PS) by incorporating gradient information into the loss function in addition to the color used by $NeRF$. While gradients are more reliable than color, they are employed to further enhance the reconstruction quality. Their choice is necessary since the neural network starts with an empty object, and using the gradient requires at least an initial surface. In our approach, the initial surface is ensured by the first part of the loss function, allowing us to use the gradient alone (without color). The second part of the loss function is provided by Equation 4.

$$\begin{aligned} \mathcal{L}_{Grad} &= \mathcal{W}_{norm}\mathcal{L}_{norm} + \mathcal{W}_{smo}\mathcal{L}_{smo} \\ Where:\ \mathcal{W}_{smo} &<< \mathcal{W}_{norm} \end{aligned} \tag{4}$$

$\mathcal{L}_{norm}$ and $\mathcal{L}_{smo}$ are provided by equations 5 and 6. $norm_{pred}$ is directly determined from the occupancy gradient provided by the network model, while $norm_{ps}$ is the normal vector obtained using a single-view photometric reconstruction technique. In our case, we use the technique proposed by Zheng et al. [14]. The MAE (Mean Angular Error) represents the angle between the two vectors, measured in radians. For continuous and smooth surfaces, a smoothing function can be added to the loss function (see Equation 6) to ensure that neighboring vectors remain close to each other. This function has a very low weight; otherwise, it would result in flat surfaces where all vectors are identical. $norm_{neig}$ is generated by a slight offset of the $norm_{pred}$ vector.

$$\mathcal{L}_{norm} = \sum_{p \in P_{surf}} MAE(norm_{pred}, norm_{ps}) \tag{5}$$

$$\mathcal{L}_{smo} = \sum_{p \in P_{surf}} \|norm_{pred} - norm_{neig}\|_2 \tag{6}$$

The final loss function is presented in Equation 7, which combines both the Silhouette and Gradient components.

$$\begin{aligned} \mathcal{L} = \mathcal{W}_{in}\mathcal{L}_{in} + \mathcal{W}_{out}\mathcal{L}_{out} + \mathcal{W}_{norm}\mathcal{L}_{norm} + \\ \mathcal{W}_{smo}\mathcal{L}_{smo} \end{aligned} \tag{7}$$

The key question is how to determine which points belong to the surface, which are inside the object, and which are outside based on the model. Most existing techniques calculate these points by starting with the image pixels. These pixels are projected along the projection rays, and a subsequent search operation is used to classify the points based on the occupancy predicted by the model. This method is slow and resource-intensive. To address this issue, we use a method proposed in [10], which involves shifting the points along the rays in the direction of the camera.

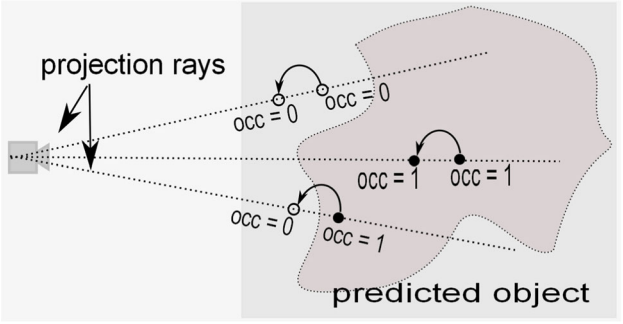


Fig. 3 Determination of points near and on the surface

When we move a point along the ray, we observe three possible cases as shown in Figure 3 :

- The point is regarded as occupied both prior to and following the displacement, thereby confirming its inclusion within the object.
- The point is unoccupied both before and after the displacement, it is considered outside the object.
- The final case is the most important, where the point is occupied before the displacement but unoccupied afterward, indicating that it is close to the surface when the displacement distance is small. We apply the 'Regula Falsi method' to determine the points that lie exactly on the surface.

We utilized the following displacement equation (Equation 8) to identify points near the surface.

$$P' = P - \delta d \cdot Ray \tag{8}$$

Here, P represents the initial point, P' represents the position after displacement, δd denotes the displacement distance, and Ray represents the projection radius.

A high value for $\mathcal{W}_{in}$ given by equation 6 compels the model to construct the SFS object, while small value ensures the correction of the surface orientation. The value of the gradient is also important, as we assume that the normals generated by Zheng et al. method [14] are correct. However, the internal occupancy provided by the SFS is inaccurate for the concave parts of the object, as shown in Figure 1. This value should have a low weight to prevent shifting the surface outward from the SFS object. We can have three scenarios for the weight of the internal occupancy $\mathcal{W}_{in}$ of the object (Figure 4):

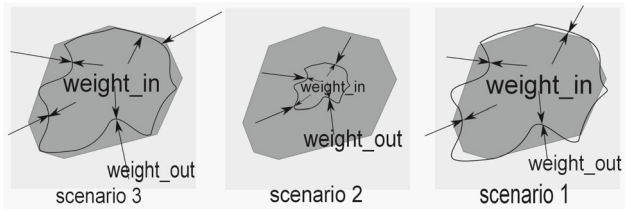


Fig. 4 Impact of intra- and inter-surface occupancy points

- Scenario 1: $\mathcal{W}_{in}$ is much higher than $\mathcal{W}_{out}$; in this case, the surface will shift toward the exterior.
- Scenario 2: $\mathcal{W}_{in}$ is significantly lower than $\mathcal{W}_{out}$; in this case, the surface will not be generated, and the model will return an occupancy of zero everywhere.
- Scenario 3: $\mathcal{W}_{in}$ is relatively low compared to $\mathcal{W}_{out}$; but it drives the surface towards the outer region, in this case, the surface is generated correctly and remains within the occupied region of the SFS.

In order to accelerate the model's convergence towards an initial object generated by the SFS data, we assign a relatively high value to $\mathcal{W}_{in}$ initially and then this value will be reduced according to the following exponential function:

$$\mathcal{W}_{in} = minValue + 2^{-3(iter/maxIter)} \tag{9}$$

This equation assigns a value to the weight $\mathcal{W}_{in}$ that ranges from 1 (similar to $\mathcal{W}_{out}$) to a minimum value that prevents the model from converging to an empty surface. This minimum value can be set through simple experimentation.

Algorithm 1 presents the pseudocode of the proposed method for 3D reconstruction. It begins with the initialization of the space parameters and the loss function, followed by the generation of an approximate shape using the Shape-from-Silhouette (SFS) technique. The algorithm then selects a set of random points from an initially generated grid. These points are displaced by a distance (δd) according to equation (8) (see also Figure 4). When a point approaches the surface, it is used to determine the intersection with the surface using the Regula Falsi method. Next, the occupancy of these points is compared with that obtained from the SFS method, while the normals of the surface points are compared to those derived from photometric stereo.

Algorithm 1 SPSN-MVPS Algorithm

```
Initialize :
  W_out, W_norm, W_smo
while nb_iter < max_iter do
  Randomly select nb_pts_batch points.
  Calculate the SFS occupancy (occ_sfs) for each point.
  for all view in train views do
    Compute the predicted occupancy (occ_pred).
    for all points with occ_pred = 1 do
      Move these points by δd according Equation 8
      if new occupancy is equal to 0 then
        Apply the Regula Falsi method to determine the surface
        points pts_surf.
        Calculate the normals (gradients) of the points pts_surf.
        Shift pts_surf to obtain a neighboring points pts_neig
        and calculate their gradients.
      end if
    end for
    Compute W_in using Equation 9
    Compute loss_value using the loss function specified by
    Equation 7.
    Perform a backward in the model and update the network
    weights.
    nb_iter++
  end for
end while
end.
```

5 Experiments

In this section, we will present the experimental results of our approach.

We used a machine equipped with an Intel i9-11900K CPU, 100GB of RAM, and an RTX A4000 GPU with 16GB of memory. We conducted our tests on the DiligenT-MV dataset [15], which includes 5 real objects captured from 10 different camera positions with 96 distinct light sources. The normal maps for each view were generated using Zheng et al. technique [14]. Our MLP architecture is similar to that of Yariv et al. [16]; we use a 8-layer MLP with a Softplus activation function and a hidden dimension of 256. (For more details, see [16]). Regarding the parameters of our loss function, which represent the core of our work, we set the weights of the different components as follows : $\mathcal{W}_{in}$ according to the equation 9 with a minimum value is 0.005, $\mathcal{W}_{out} = 1.0$, $\mathcal{W}_{norm} = 1.0$, $\mathcal{W}_{smo} = 0.005$.

For the algorithm parameters, we have the following configuration: The point displacement distance is set to 0.4 (according to the reconstructed space). Learning rate : 0.001. The reconstruction space is divided into a 300x300x300 grid, with values ranging from -2 to 2. The number of points per batch is set to 120,000, this parameter is machine-dependent and may vary based on the hardware used. These parameters were determined through extensive testing. We note that the proposed approach does not require the minimum and maximum distances (such as [8, 17]) between the object and the camera, as we rely on the projection of 3D points onto the images. However, for image visualization, we used the same method as PSNerf [12], though this phase is optional.

We begin our evaluation with an illustrative example of our model's convergence, applied to the cow object using the proposed loss function. Figure 5 shows the evolution of the

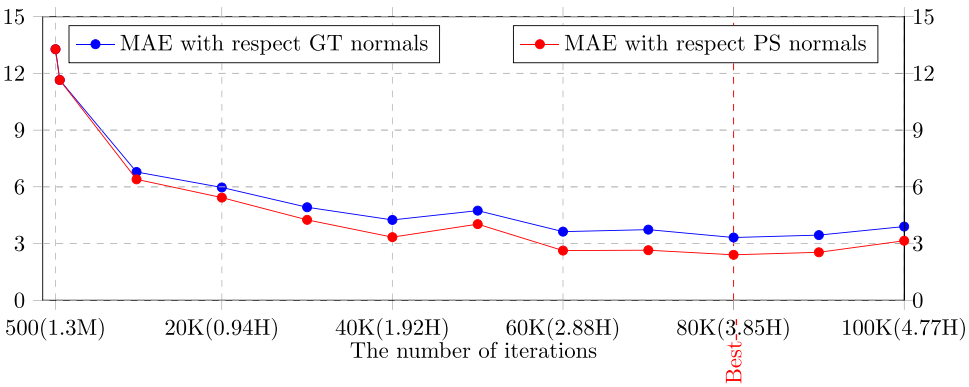


Fig. 5 The MAE with respect to the number of iterations for "Cow" object

Mean Angular Error (MAE, in degrees) over the course of iterations, along with the associated processing time.

The blue curve represents the model's convergence with respect to the normals estimated via the Photometric Stereo (PS) method, while the red curve corresponds to the ground truth normals.

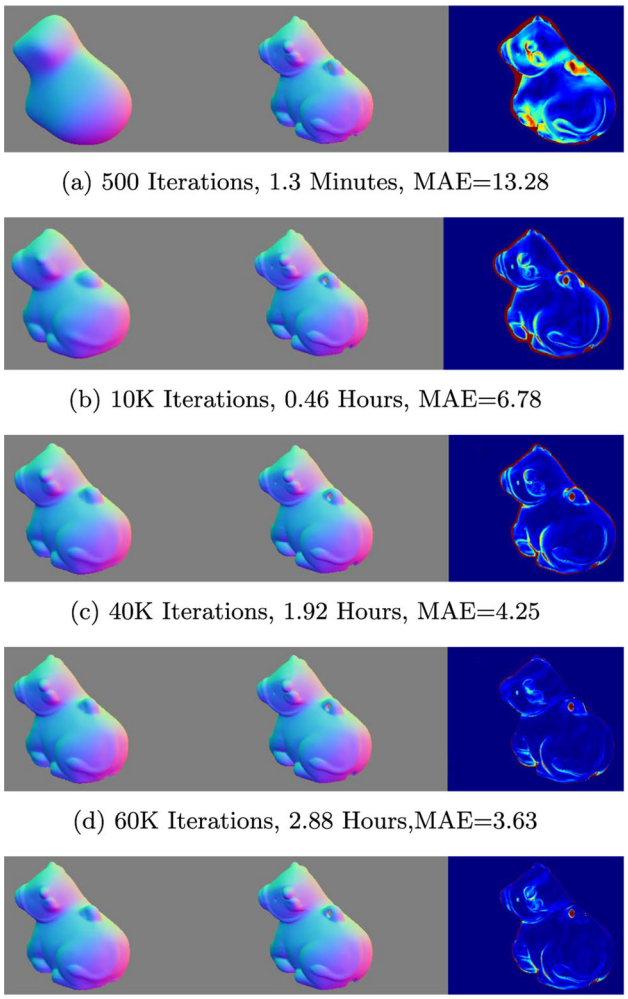
We observe that both curves tend toward low minimum values, indicating good convergence. However, due to the random selection of grid points at each iteration, a slight instability can be observed in the convergence path. Despite this, both curves follow a similar trend, allowing us to estimate the minimum value associated with the (unknown) ground truth normals based on the minimum reached by the (known) PS normals.

These results are very promising, particularly in terms of processing time. Our method requires approximately 17 seconds per 100 iterations. The model typically converges around 90,000 iterations, for example, the cow object converges at around 80,000 iterations, and the bear object at 70,000. This allows us to significantly reduce the computation time to approximately 3 hours, compared to 12 hours in the case of PS-NeRF [12].

Figure 6 illustrates the improvement of the gradient for the "cow" object with respect to the number of iterations and execution time. The first subfigure (a) shows that the object initially covers the entire area occupied by the silhouettes, exactly as depicted in Figure 2, meaning that the internal occupancy of the SFS is very high. Subsequently, this occupancy is reduced, and the surface quality is improved by decreasing the weight of the internal occupancy.

Figure 7 illustrates the evolution of the 3D object "cow" with respect to the number of iterations and processing time. This figure visually demonstrate the quality of the objects generated by the proposed method. The convergence is fast, especially in the first few minutes, toward the initial object generated by the shape-from-silhouette method.

A relatively good object is obtained in approximately 30 minutes, which represents a significant advantage of the proposed method. These objects can be used in applications where computational speed is prioritized over object quality.



(a) 500 Iterations, 1.3 Minutes, MAE=13.28

(b) 10K Iterations, 0.46 Hours, MAE=6.78

(c) 40K Iterations, 1.92 Hours, MAE=4.25

(d) 60K Iterations, 2.88 Hours,MAE=3.63

(e) 80K Iterations, 3.85 Hours,MAE=3.32

Fig. 6 The evolution of the gradient with respect to the number of iterations for "Cow" object

Figure 8 shows the gradient of other objects (Bear, Pot2, Buddha, and Reading) at the best iterations achieved using

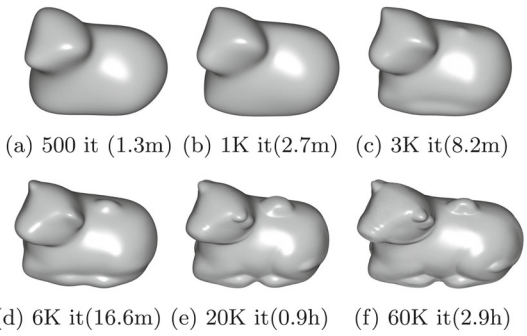


Fig. 7 Evolution of 3D objects for "Cow" object

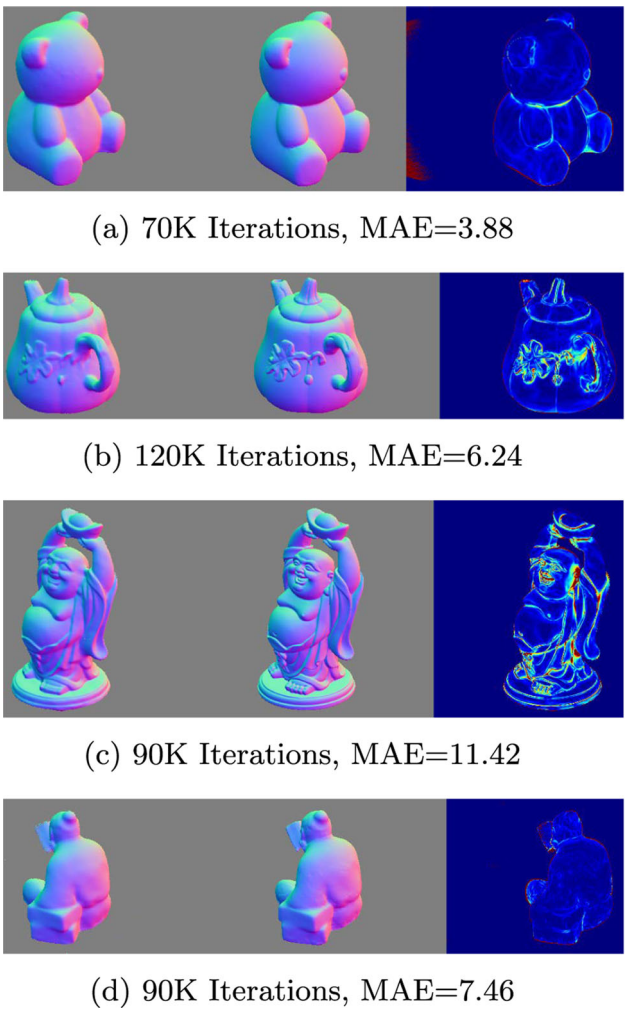


Fig. 8 Normal vectors (Gradient) of Bear, Pot2 and Buddha objects

Fig. 9 3D objects, for "Bear", "Pot2" and "Buddha" objects

Table 1 Comparison with recent techniques

DiliGenT-MV (Mean Angular Error)					
Method	Bear	Buddha	Cow	Pot2	Reading
NeRFactor [18]	12.68	25.71	17.87	15.46	21.24
NeRD [19]	19.49	30.41	33.18	28.16	30.83
PhySG [18]	11.22	26.31	11.53	13.74	25.74
Nerf^{+N} [12]	7.03	13.50	8.26	7.93	14.01
Unisurf^{+N} [12]	4.26	11.29	5.05	6.37	9.58
MVAS [17]	**3.08**	**9.90**	3.72	**5.07**	10.02
PSNERF [12]	3.21	10.10	4.08	5.67	8.83
SPSN(Ours)	3.88	11.42	**3.32**	6.24	**7.46**

the proposed method. The results demonstrate a significant improvement in quality.

In summary, Table 1 presents a quantitative evaluation of the proposed technique using the Mean Angular Error (MAE), computed on five objects from the Diligent dataset. This metric measures the accuracy of the estimated normals with respect to the ground truth normals, and is a standard criterion in the field of photometric reconstruction. We compare our method to several reference approaches belonging to the same category. The results show that our method achieves the best performance for the cow and reading objects, outperforming all the techniques listed, as previously discussed. In addition to this increased accuracy, our approach also offers a significantly reduced computation time, which strengthens its relevance for applications requiring both efficiency and robustness. For the other objects (ball, cat, pot), although our method does not achieve the best MAE, the performance remains close to that of the best-known methods, demonstrating good generalization of our approach across different geometric shapes.

6 Discussion

In the previous section, we conducted a comparison with recent and well-known techniques from the literature. Most inverse rendering techniques rely on image pixels to determine their projections onto the surface, which requires extensive search processes. Moreover, all points used to train the model depend solely on the pixel rays, and the mask verification is performed only from the projection view, even if the point does not belong to other views which requires a long processing time and prolonged convergence.

In our method, we propose incorporating the concept of SFS (Shape-from-Silhouette) occupancy, which is generated simultaneously from all views. This helps the model converge faster with an initial object that is much closer to the real object, unlike PS-NeRF ([12]), which starts from an empty space, or MVAS [17], which begins with a sphere. Notably, SFS occupancy has the same nature as the output of our model, unlike RGB values or gradients, which contribute to slower convergence.

Using spatial points as training points allows the entire space to be trained simultaneously, including both occupied and unoccupied points from the SFS, in addition to surface points guided by the gradient. This innovative approach can significantly enhance inverse rendering-based reconstruction methods. However, further improvements are needed to achieve even better results.

7 Conclusion

In this paper, we addressed the challenges related to processing time and convergence in 3D reconstruction by guiding the model using an initial object generated from the Shape-from-Silhouette (SFS) method. This strategy significantly reduces computation time while improving reconstruction efficiency. By incorporating SFS information into the loss function, formulated from 3D space points, our approach promotes more stable and reliable convergence compared to methods that rely solely on color and normal data. It also eliminates the need to manually define Near and Far distances, as these are automatically determined by the SFS algorithm, making the reconstruction process both faster and more robust.

Despite these advancements, some limitations remain, particularly in reconstructing objects with holes. Certain methods aim to enhance surface quality by refining normal estimation, which is a key component in multi-view reconstruction. However, such normal-based approaches often depend on controlled lighting conditions, which limits their practical applicability. Future work will focus on improving normal estimation and integrating additional information into the loss function to further enhance reconstruction quality while reducing dependence on specific lighting setups.

Author Contributions All authors contributed to the discussion of the technique and/or the revision of the manuscript.

Data Availability DiLiGenT-MV dataset is available at: https://sites.google.com/site/photometricstereodata/mv.

Declarations

Competing interests The authors declare no competing interests.

References

1. Deng, L., Chen, S., Ji, Y., Wang, J., Yang, X., Huang, S.: 3d synthetic ct patch generation and reconstruction by using multi-resolution generative adversarial network. Signal, Image and Video Processing **19**(3), 267 (2025)
2. Li, Y., Xia, H., Liu, Y., Sun, Q., Huo, L., Ni, X.: Research on the detection method of phenotypic information of pinus massoniana lamb. seedling root system. Signal, Image and Video Processing **18**(10), 6961–6972 (2024)
3. Lan, T., Yang, G.: A low-cost pipeline surface 3d detection method used on robots. Signal, Image and Video Processing **18**(4), 3915–3924 (2024)
4. Liu, M., Wang, W., Zhao, W.: Pva-gcn: point-voxel absorbing graph convolutional network for 3d human pose estimation from monocular video. Signal, Image and Video Processing **18**(4), 3627–3641 (2024)
5. Shan, L., Sun, J., Hong, B., Kong, M.: 3d reconstruction of flame temperature field based on lightweight residual network with spatial attention mechanism. Signal, Image and Video Processing **18**(10), 6661–6670 (2024)
6. Khrouch, H., Mahdaoui, A., Marhraoui Hsaini, A., Merras, M., Chana, I., Bouazi, A.: Improving camera parameter estimation using an adaptive genetic algorithm. Signal, Image and Video Processing **19**(1), 1–15 (2025)
7. Abada, L., Aouat, S.: Improved photometric stereo based on local search. Multimedia Tools and Applications **81**(21), 31181–31195 (2022)
8. Gao, K., Gao, Y., He, H., Lu, D., Xu, L., Li, J.: Nerf: Neural radiance field in 3d vision, a comprehensive review. arXiv preprint arXiv:2210.00379 (2022)
9. Oechsle, M., Peng, S., Geiger, A.: Unisurf: Unifying neural implicit surfaces and radiance fields for multi-view reconstruction. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 5589–5599 (2021)
10. Abada, L., Mezabiat, A.S., Gacem, T., Malki, O.C., Mekkaoui, M.: Enhancing psnerf with shape-from-silhouette for efficient and accurate 3d reconstruction. Multimedia Tools and Applications, 1–15 (2024)
11. Abada, L., Bennaceur, M., Boudjenana, A.A., Aouat, S.: Using pso metaheuristic to solve photometric 3d reconstruction. In: 2022 7th International Conference on Image and Signal Processing and Their Applications (ISPA), pp. 1–6 (2022). IEEE
12. Yang, W., Chen, G., Chen, C., Chen, Z., Wong, K.-Y.K.: Ps-nerf: Neural inverse rendering for multi-view photometric stereo. In: European Conference on Computer Vision, pp. 266–284 (2022). Springer

13. Cheung, K.-M., Baker, S., Kanade, T.: Shape-from-silhouette across time part i: Theory and algorithms. International Journal of Computer Vision **62**, 221–247 (2005)
14. Zheng, Q., Kumar, A., Shi, B., Pan, G.: Numerical reflectance compensation for non-lambertian photometric stereo. IEEE Transactions on Image Processing **28**(7), 3177–3191 (2019)
15. Li, M., Zhou, Z., Wu, Z., Shi, B., Diao, C., Tan, P.: Multi-view photometric stereo: A robust solution and benchmark dataset for spatially varying isotropic materials. IEEE Transactions on Image Processing **29**, 4159–4173 (2020). https://doi.org/10.1109/TIP.2020.2968818
16. Yariv, L., Kasten, Y., Moran, D., Galun, M., Atzmon, M., Ronen, B., Lipman, Y.: Multiview neural surface reconstruction by disentangling geometry and appearance. Advances in Neural Information Processing Systems **33**, 2492–2502 (2020)
17. Cao, X., Santo, H., Okura, F., Matsushita, Y.: Multi-view azimuth stereo via tangent space consistency. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 825–834 (2023)
18. Zhang, K., Luan, F., Wang, Q., Bala, K., Snavely, N.: Physg: Inverse rendering with spherical gaussians for physics-based material editing and relighting. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5453–5462 (2021)
19. Boss, M., Jampani, V., Braun, R., Liu, C., Barron, J., Lensch, H.: Neural-pil: Neural pre-integrated lighting for reflectance decomposition. Advances in Neural Information Processing Systems **34**, 10691–10704 (2021)