Estimation of 3D Shape and Volume of Fire Plumes from Multiple Views

J A Blanco[1], M Pardàs[1], J R Casas[1], R Paugam[2] A Àgueda[2], J Wagner[3], R A Parsons[4] and E Planas[2*]

[1] Image Processing Group (GPI), Department of Signal Theory and Communications, Universitat Politècnica de Catalunya·BarcelonaTech, Jordi Girona 1-3, Edifici D5, 08034 Barcelona, Spain
[2] Centre for Technological Risk Studies (CERTEC), Chemical Engineering Department, Universitat Politècnica de Catalunya·BarcelonaTech, Av. Eduard Maristany, 16, 08019 Barcelona, Spain
[3] University of Montana. 406 ISB, 32 Campus Dr. Missoula, Mt 59812.
[4] Fire Sciences Laboratory, Rocky Mountain Research Station, US Forest Service, 5775 W. Highway 10, Missoula, MT 59801, USA.

*Corresponding author: eulalia.planas@upc.edu

Abstract. This study evaluates deep-learning and Shape from Silhouette (SfS) methods for 3D reconstruction of smoke plumes. It demonstrates the deep-learning method's superiority in cases with limited camera views and calibration data, achieving high-quality reconstructions of semi-transparent smoke without precise calibration. The research emphasizes the significance of pre-processing and data appearance for neural network efficacy. By improving 3D reconstruction techniques, this work aids in advancing wildfire tracking and environmental analysis, offering a practical approach for real-world applications in fire science.

1. Introduction

To understand and forecast wildfire behavior and their associated smoke plumes, physics-based models are often used. However, they must be previously validated with empirical datasets. Gathering data on the progression of wildfire smoke plume shapes is notably challenging due to the plumes' complex 3D dynamics and the optical properties involved. While active remote sensing tools like LIDAR [1]—which are both costly and specialized—can collect such data, capturing still images or videos of the smoke plumes from various angles offers a more accessible alternative. This method, when combined with advanced computer vision techniques, may enable the detailed mapping of the plumes in three dimensions (3D).

Current methods for 3D reconstruction often depend on texture cues, but these methods face substantial challenges with semi-transparent or feature-sparse surfaces, such as smoke. To reconstruct smoke plumes in 3D, two primary strategies are employed. The first involves a direct, single-step 3D reconstruction using a Structure from Motion (SfM) algorithm. Notable examples include Gomez et al. [2], who successfully applied SfM software to a volcano's smoke plume, and De Donato et al. [3] and Hu et al. [4], who achieved 3D reconstructions of stack emissions and gas clouds using geometric methods, respectively. The alternative strategy initiates with segmenting the target, followed by volume computation. Attempts have been made previously to reconstruct volumes using Shape from Silhouette

(SfS) [5] utilizing segmentation masks. Both methods, however, necessitate accurately determined camera parameters.

A critical factor in this context is the number of viewpoints available for creating a 3D reconstruction. The references provided before have over 4 views available, with figures sometimes reaching up to fifty. Some advancements were recently made with fewer views, including extreme scenarios of single-view reconstructions [6]. In wildfire datasets, typically less than four views are available, complicating the 3D reconstruction of smoke plumes further.

This paper explores two approaches to address these challenges based on the available resources and number of views: a traditional SfS technique and a deep learning method. We evaluated qualitatively and quantitatively both methods, using real and synthetic images.

2. Methodology

Two distinct methodologies were employed for the reconstruction of smoke plumes: a voxel-based SfS technique and a deep learning approach utilizing Kim’s Sketch to Fluid model [7]. To facilitate the latter, a preprocessing pipeline was developed, which convert the images to sketches, to ensure the input data matched the training domain of the network. A notable distinction between these methodologies lies in their treatment of spatial data. The SfS approach incorporates precise world position and scale information for the plume, whereas the deep learning method outputs the reconstructed surface without inherent positional or scale details. However, when camera data is available, as in the SfS method, the reconstructed object's position and scale can be accurately determined and adjusted using the bounding box dimensions derived from the SfS analysis.

2.1. Voxel-based SfS

A baseline Matlab implementation [8] was used to compute the visual hull and to assess the performance of SfS. The necessary inputs are *n* silhouettes and *n* camera matrices. The camera matrix contains the intrinsic (focal length, optical center, skew) and extrinsic information (position and rotation) of the camera. This additional information is what allows the algorithm to get a reconstruction that matches the position, rotation and scale of the real object. Given that classic SfS approaches require the intrinsic and extrinsic camera parameters, in wildfire scenarios this Voxel-based SfS approach would require camera calibration as a necessary step before carrying out the reconstruction. As shown in Fig. 1, this module can be divided into three different stages: bounding box estimation, visual hull estimation and mesh conversion.

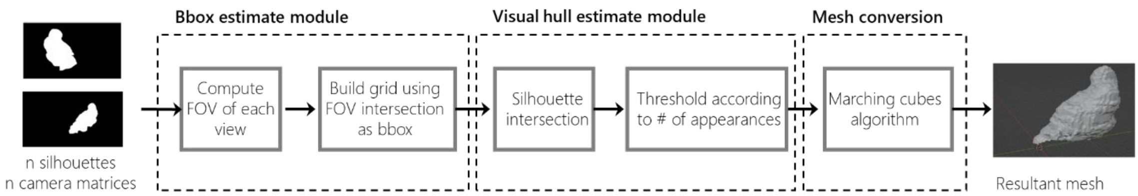

Figure 1. Overview of the SfS method

Bounding box estimation: First of all, the relative size and position of the object have to be estimated, that is, its bounding box in the 3D space. This is obtained by intersecting the fields of view of all the given cameras.

Visual hull estimation: Each silhouette pixel is projected from the respective view using the camera matrix. The grid created is then thresholded to discern object-containing voxels with the threshold set just below the total number of views, thereby requiring a voxel to intersect with nearly all silhouette projections to be considered smoke. This is a rather strict criteria, since if a part of the smoke is not detected in more than 2 views, it would not appear in the reconstruction. In real scenarios this threshold can be adapted to be more tolerant.

Mesh conversion estimation: The last step is to convert the voxels into a format that can be loaded in 3D software. In this case, the voxel grid after being thresholded is converted to an isosurface. The conversion from voxel to mesh is done using the marching cubes algorithm [9].

2.2. Sketch to fluid algorithm

Kim's Convolutional Neural Network (CNN) deep-learning approach aims to use sketches from artists as input, instead of real images. It has been adapted for the case of wildfire smoke plume images with a pre-processing and a post-processing module. An overview of the system can be seen in Fig. 2

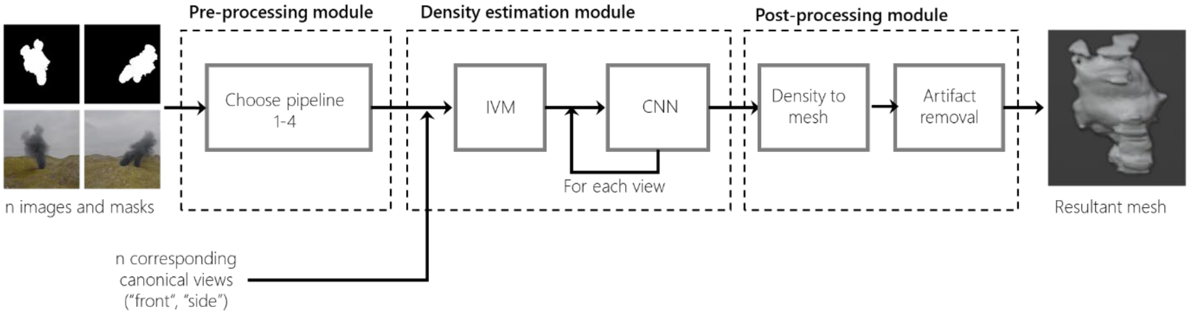

Figure 2. Overview of the proposed adaptation of Kim's Sketch to Fluid model with the pre- and post-processing modules. IVM: Initial Volume Modelling.

Pre-processing module: Four different pipelines of preprocessing have been tested to assess the system with real and Blender-simulated images (Fig. 3). As the sketcher module images look consistent and have a flat look, the preprocessing methods rely on quantizing the image with different levels and gray values. After the pre-processing stage, the images are input to the sketch to fluid module.

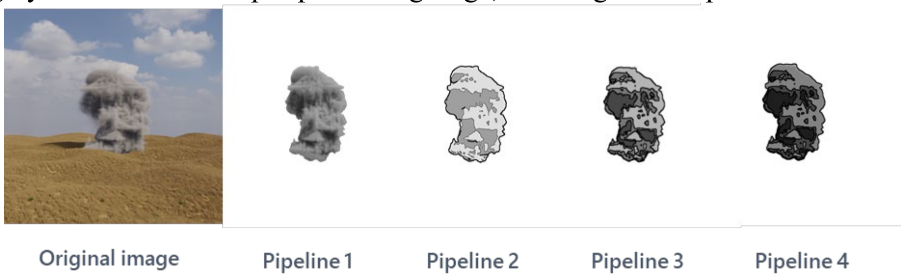

Figure 3. Synthetized smoke Blender image (left) and sketch outputs obtained using the diverse pipelines used.

Density estimation module: To obtain a density from these inputs several steps are required. First, a series of canonical views are defined which correspond to the orientation that the input images have been taken. Afterwards, an initial density using the silhouettes of 2 views (side, front) is obtained using the IVM (Initial Volume Modeling) module. This consists in using the mask inputs of the different viewpoints and computing a visual hull from the intersection of these silhouettes projected from the side and front view. In this step it is assumed that images are obtained with parallel projection and the views are orthogonal. In practice, it restricts the images to be orthogonal views acquired at the same distance from the plume. Finally, Gaussian blur is applied to smooth the boundaries. This density is refined later by the CNN using an iterative approach where for each image-view pair given, the reconstruction is input to the CNN.

Post-processing module: The outputs of the sketch to fluid CNN are not meshes, but rather a grid with the density of the plume in each position. Because of that, a post-processing pipeline is needed to convert it to a mesh, as well as to remove any possible artifacts. This process is done within Blender [10], which provides a Python library that allows carrying out both steps.

2.3. Experimental data

To assess the performance of the two methodologies proposed to obtain the 3D fire plume reconstruction, two type of data were used: images corresponding to real smoke plumes and images corresponding to smoke plumes synthesized using Blender.

Two subsets of real images were utilized. One with 15 single-view images of real wildfires [11] (results not shown) and a video sequence of artificial smoke released in a laboratory with front and side views (Fig. 4). Due to the absence of actual volume data or pre-existing 3D reconstructions of the smoke, this evaluation was conducted qualitatively. Furthermore, the lack of calibration data for these image sets restricted the assessment to only the deep-learning approach when working with real images.

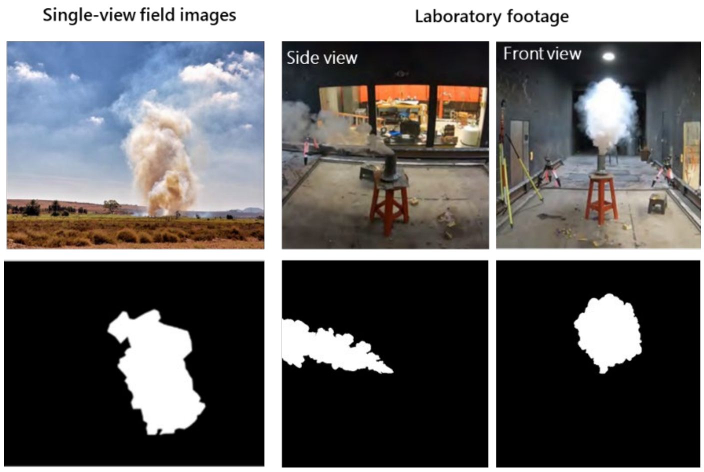

Figure 4. Example of the real images used (top) with the ground truth masks of the smoke (bottom).

The quantitative assessment of the Shape from Silhouette (SfS) algorithm's effectiveness requires a ground truth for the target's 3D shape. Synthetic simulations of smoke plumes, created in Blender [10], were used for this purpose. This approach provides a controlled virtual environment where camera placement is flexible, and both intrinsic and extrinsic parameters are readily accessible. Blender's simulations offer a volumetric representation of smoke plumes, which serves as the ground truth for evaluating the reconstruction methods. These simulations depict outdoor scenes with time-based modeling of the physical behavior of various smoke plumes. (Fig. 5).

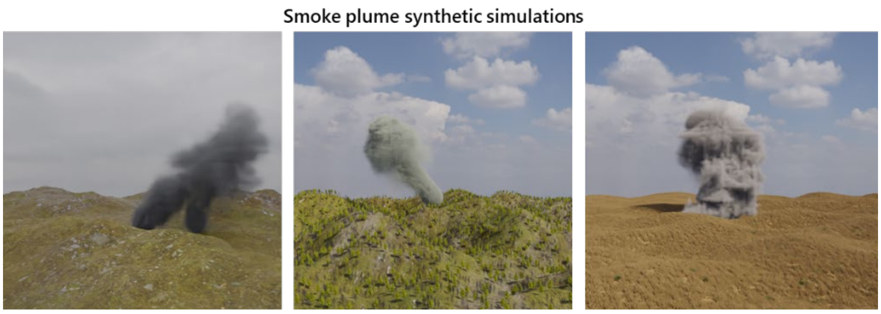

Figure 5. Shots of the three different synthetic simulations used.

3. Results

Several metrics were employed to assess the quality of the reconstructions. Specifically, the total volume of the reconstructed smoke plume and 3D Intersection over Union (IoU) were used for volumetric comparison. The Hausdorff and Chamfer distances were applied to evaluate the surface accuracy of the reconstructed shape. For qualitative analysis, images showcasing the reconstructions from various viewpoints were also utilized.

3.1. Synthetic data results

3.1.1. Pre-processing assessment. A series of experiments across three simulations was conducted to identify the optimal pre-processing strategy for applying the deep-learning approach to real images. The outcomes of each pre-processing method were recorded and compared against the SfS technique. For both approaches, a single front and side view were used, with symmetry assumed to simulate four views.

The results are shown in Table 1. The second preprocessing pipeline emerged as the most consistent, surpassing the SfS method across all metrics. Employing the CNN approach yielded smoother outcomes than SfS with an equivalent number of views. When limited views are available, the SfS method tends to produce reconstructions with noticeable artifacts due to camera frustums, resulting in a box-like 3D shape. Given its superior performance both qualitatively and quantitatively, the second preprocessing pipeline has been chosen for all subsequent assessment tests.

Table 1. Average results of SfS and deep-learning CNN approach with the 4 different pre-processing methods.

	Volume relative error	IoU	Hausdorff distance	Chamfer distance
SfS	0.349	0.620	4.27	1.56
Pre-processing pipeline 1	0.155	0.676	3.65	1.40
Pre-processing pipeline 2	0.141	0.719	3.20	1.07
Pre-processing pipeline 3	0.052	0.639	4.03	1.18
Pre-processing pipeline 4	0.319	0.630	14.08	4.36

3.1.2. Impact of available views on reconstruction quality. Two types of evaluations were conducted. Firstly, the performance of the CNN approach was assessed with one, two, or four views. For the four-view scenario, only two original views were generated from Blender, with the remaining two created by flipping the original images. This technique aims to more accurately reflect real-world wildfire situations, where obtaining four simultaneous views is challenging. Secondly, the SfS method was evaluated using up to thirty-two views, plus an additional top camera view.

The outcomes from the CNN approach highlight that the choice of view for a single-view input significantly influences the results, varying with the selected viewpoint. Additionally, given that smoke plumes often exhibit symmetry, assuming symmetry tends to enhance the results, particularly when observed from a top-view perspective.

The impact of adding more cameras on the Shape from Silhouette algorithm's performance is depicted in Fig. 6, where the reconstruction was also evaluated with an additional top view. The inclusion of a top view significantly enhances the reconstruction outcomes across all metrics, though its effectiveness diminishes as the number of available cameras increases. Moreover, while augmenting the number of cameras generally improves results, the improvement plateaus beyond eight cameras, showing minimal variation with further additions.

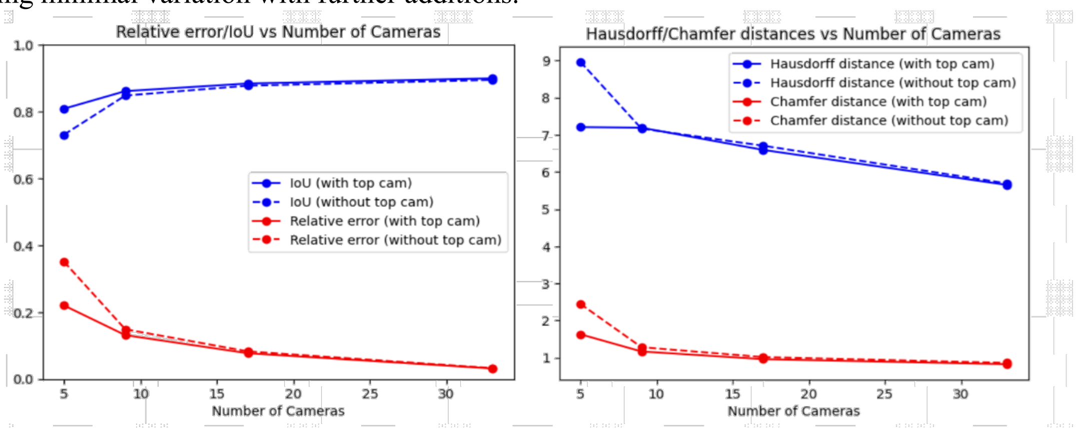

Figure 6. Results of the SfS when increasing the number of cameras, with and without a top camera.

3.2. Real data results with deep learning approach

A temporal series of volumes was reconstructed from the laboratory artificial smoke video sequence, evaluating four frames spaced 10 seconds apart. Each frame was assessed using a side and a front view. This experiment posed a significant challenge, as the smoke was barely visible from the side view without distinct shadows. Despite this, the reconstructions closely matched the original smoke plumes, as illustrated in Fig. 7. The reconstructions from both front and side views closely approximated the actual data, despite cameras were not set completely perpendicular, and the temporal progression remained consistent. This demonstrates that reconstructions of less visible smoke are feasible through the pre-processing method, provided the masks are accurately detected.

Figure 7. Comparison of the original footage (top row) and reconstruction (bottom row) over 4 different frames of the laboratory experiment (front and side views).

4. Conclusions

This study investigated 3D reconstruction of smoke plumes using silhouette masks and multi-view imagery, comparing a deep-learning algorithm with the conventional Shape from Silhouette technique. Findings reveal deep-learning's advantage, especially under limited camera views and calibration data, achieving precise reconstructions as long as orthogonal views from a similar distance to the plume are available. Quantitative evaluations of both methods underscore the enhanced accuracy of 3D reconstructions with deep-learning. Moreover, the study highlights the crucial role of data appearance and pre-processing in neural network reconstructions, presenting a new approach for effectively modeling semi-transparent objects like smoke plumes.

Acknowledgements

This research was funded by the Strategic Environmental Research and Development Program (SERDP), grant number RC20-1025. Funding also provided by Spanish Research Agency, grant number PID2020-114766RB-100 of MCIN/AEI/10.13039/501100011033. A. Àgueda is a Serra Hunter fellow.

References

[1] Lareau N P and Clements C B 2017. *J. Appl. Meteor. Climatol.* **56** 2289–2299
[2] Gomez C and Kennedy B 2018 *J. Volcanol. Geotherm. Res.* **350** 84–88
[3] De Donato P, Barres O, Sausse J, Martin D 2018 *Remote Sens*. **10**, 678
[4] Hu Y, Xu L, Xu H et al. 2022. *Opt. Express* **30** 25581–25596
[5] Baumgart B G 1974. Stanford University report STAN-CS-74-463 https://apps.dtic.mil/sti/pdfs/ADA002261.pdf
[6] Yang X, Lin G and Zhou L 2023 *IEEE Trans. Image Process.***32** 3746-3758
[7] Kim B, Huang X, Wuelfroth L, Tang J, Cordonnier G, Gross M and Solenthaler B 2022. *Computer Graphics Forum* **41**, 97–110
[8] Mikhnevich M 2013 Visual hull implementation. https://github.com/maximm8/VisualHull
[9] Lorensen W E and Cline H E 1987 Marching cubes: A high resolution 3D surface construction algorithm *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques* 163–169
[10] Blender Online Community 2018 Blender - A 3D modelling and rendering package. Blender Foundation, Stichting Blender Foundation, Amsterdam.
[11] Baldrich A 2023 Smoke Plume Segmentation of Wildfire Images *Master's Thesis*, UPC, Escola d'Enginyeria de Barcelona Est, Departament d'Enginyeria Química.