INTERACTIVE SEGMENTATION FOR SHAPE FROM SHADING OVER HR SAR IMAGES

Franco Marchesoni-Acland [†] *Marie d'Autume*[†] *Gabriele Facciolo*[†]
Carlo de Franchis[†‡] *Jean-Michel Morel*[†] *Enric Meinhardt-Llopis*[†]

[†]Centre Borelli, ENS Paris-Saclay, Université Paris-Saclay, [‡] Kayrros SAS

ABSTRACT

Shape from shading (SfS) enables 3D reconstruction of stockpiles from a single image. However, this method requires proper boundary conditions to work properly. Obtaining such Dirichlet and Neumann conditions is equivalent to a segmentation of the heaps. To get a fast and accurate 3D reconstruction, we propose a simple and interactive segmentation method. SfS is then applied on 0.5-meter resolution Synthetic Aperture Radar (SAR) images with more precise boundary conditions. The results show that prior segmentation is preferable to no segmentation for the stockpiles volume estimation problem. Furthermore, we show that the proposed interactive segmentation method reduces the annotation time needed for such a prior segmentation

***Index Terms*—** Synthetic Aperture Radar, Shape-from-shading, interactive segmentation, Dirichlet-Neumann boundary conditions

1. INTRODUCTION

The storage and management of stockpiles of different materials is fundamental to many industries, such as mining, construction, or waste management. This monitoring was done manually with on-site surveys [1], but this approach is difficult and costly. Remote sensing tools like photography from planes or Unmanned Aerial Vehicles (UAVs) are often preferred to site monitoring [2, 3]. Satellite images provide a high revisit time and worldwide coverage [4]. SAR imaging does not require sunlight and is free from the influence of atmospheric conditions.

Shape-from-Shading (SfS) [5] estimates the volume of a stockpile from a single image. It requires the scene to be composed of material with homogeneous reflective properties. The region of interest must be carefully defined to ensure this hypothesis.

In this work, we present an interactive segmentation tool as a prior step to shape from shading. To illustrate this approach, we tackle the problem of coal volume estimation from high-resolution SAR imaging. This case has two big differences with the work in [4]: in the current paper the images are higher resolution SAR images and we introduce an interactive segmentation approach to provide mask priors. In this particular problem, coal stockpiles are located near harbors. The sites have cranes on rails moving over the stockpiles adding and removing material. This problem is illustrative because, as in many others in which annotations are scarce and sites differ, no segmentation tool works out-of-the-box.

Our method involves three stages: the first segments overlapping objects, e.g. cranes, and crops the image. The second segments the objects of interest. The third applies SfS with the obtained boundary conditions.

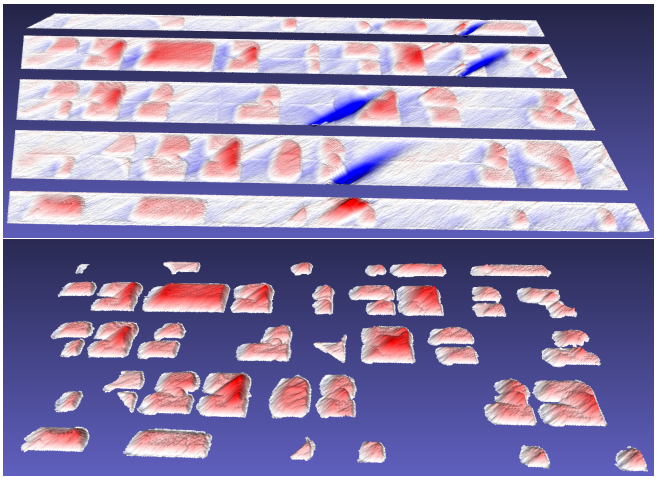

Fig. 1: 3D reconstruction using Shape-from-Shading before and after providing segmentations of the individual stockpiles.

Plan: an overview of image segmentation, interactive segmentation, and shape from shading literature is presented in Section 2. The data at hand is described in Section 3. Section 4 describes the interactive segmentation method and the SfS approach. Results and conclusions are given in Section 5 and Section 6, respectively.

2. RELATED WORK

2.1. Image segmentation

Image segmentation refers to the problem of classifying each pixel of an image. State-of-the-art methods are deep learning methods that are trained or fine-tuned over annotated datasets [6]. For problems in which there are no annotated data available, the solution is to use pre-trained models [6], unsupervised segmentation methods, e.g. [7], or to annotate the data. Transfer learning suffers when the domain shift is big, e.g. training with consumer images and applying on satellite images. Unsupervised segmentation methods traditionally perform feature extraction (classic filter stacks or neural networks [8]) and clustering [9]. When involving neural networks these methods do not necessarily need annotations, but they usually require domain data, which is not always available in large enough quantities.

2.2. Interactive segmentation

Annotations are needed in most of the cases above, to provide information relevant to general vision abilities and also to define the problem. The need for annotations is the motivation that drives the research in interactive segmentation and active learning. While active

This work was partially financed by IDEX Paris-Saclay IDI 2016, ANR-11-IDEX-0003-02, Office of Naval research grant N00014-17-1-2552, STIC-Amsud 20-STIC-04, DGA Astrid project "filmer la Terre" n° ANR-17-ASTR-0013-01, MENRT. This work was also using HPC resources from GENCI–IDRIS (grant 2021-AD011011801R1).

learning is concerned with reducing the number of annotations [10], interactive segmentation reduces their cost [11]. Better annotation efficiency can help solve big problems, e.g. in autonomous driving, medical imaging, or remote sensing. However, current annotation tools do not include state-of-the-art methods. At the same time, most state-of-the-art methods present the same complication as in supervised segmentation: good results over the benchmarks are due to the use of similar training data, many times provided in the benchmarks themselves [12]. This is why they struggle to generalize to different data, and more importantly, to quickly learn the objective of the segmentation on a new domain.

The work [13] advanced in this direction, which is in between the fields of continuous learning, few-shot learning, transfer learning, and representation learning. This line of research is still in its infancy. This is why, for many problems, specific tools have to be designed. In this paper, we describe one such tool.

2.3. Shape-from-shading

The SfS problem estimates the three-dimensional shape of a surface from one image of this surface. It was first introduced by Horn [5] and is modeled by the "image irradiance equation"

$$I(x, y) = R(\mathbf{n}(x, y)), \tag{1}$$

where $I(x, y)$ is the gray level measured at pixel (x, y) and R is the reflectance function, giving the value of the light re-emitted by the surface $z = u(x, y)$ as a function of its 3D normal $\mathbf{n}(x, y)$. One of the features that make this problem interesting is that the input data are minimal: a single gray level image is used.

Most SfS methods assume the surface as Lambertian with a constant and known albedo, a unique far enough light source so that the incident direction may be taken as constant, and a viewer standing far enough, so that the direction to the viewer is roughly constant in the scene. Under these hypotheses, the reflectance function is

$$R(x, y) = A\ \mathbf{d} \cdot \mathbf{n}, \tag{2}$$

where the vector $\mathbf{d} = (\alpha, \beta, \gamma)$ is the incident light direction and A is a constant modeling several proportionality factors, among which the albedo of the surface and the light source intensity.

Many numerical methods have been proposed for solving the SfS problem [14]. Although SAR images do not exhibit Lambertian reflectance properties, in this work we use a method based on the linearization of the reflectance function, thus we assume Lambertianity and we solve the equation in the range-azimuth plane.

3. DATA

SAR is an active sensor that first transmits microwave signals and then receives back the signals that are reflected from the surface of the Earth. Taking into account the motion of the radar antenna, the received signals are then processed to produce an image. One advantage of SAR images over optical images is that they are almost not affected by meteorological phenomena, e.g. clouds. The images we shall use were provided by Capella.

Capella operates a constellation of seven satellites, each carrying an X-band SAR. Two additional satellites are launching in 2022. The SAR instrument can be operated in spotlight, sliding spotlight, and stripmap modes. It acquires single polarization images. The images used in this work were acquired in sliding spotlight mode, covering a 5 km by 10 km area at 0.5 meter slant-range resolution, and were multi-looked, orthorectified, and geocoded on a 60 cm UTM grid (GEO product).

We worked with 6 images of the same site taken at different dates with intervals ranging from 2.5 days to 2 months. All images are cropped to a size of 2432×2337 pixels, corresponding to an area of $\sim$1.4$\times$1.4km. An example image we will use throughout the paper is Figure 2. For visualization purposes, we clipped the 3% largest values.

4. INTERACTIVE SEGMENTATION METHOD

The interactive segmentation method has three main components. The first is the determination of the region of interest. The second component is the interactive segmentation of the cranes. The last component is the interactive segmentation of the coal heaps.

The method is summarized as follows: if there is not a region of interest for the site, it is generated with polygon-based tools. The site information is then taken into account (see Subsection 4.1); cranes are removed according to Subsection 4.2; lastly, stockpiles are segmented according to Algorithm 1.

4.1. Site information

Only once for each site and in a very short time, some site information is extracted, which is useful to ignore parts of the images that are irrelevant for our use case and facilitates further processing. The main site-information is the region of interest, which is the same from image to image and whose determination amounts to a negligible annotation time. This is why it has not been mentioned in previous works [4].

More specifically, the site information consists of a rotation angle, a region of interest, and a segmentation of the rails. The rotation is conducted because the stockpile formation follows the natural direction of the rails, a diagonal in Figure 2. The angle is easily obtained by selecting two points in a rail line. The region of interest is determined by a polygon. Finally, the rails are masked out, after rotation, using bounding boxes. The angle, mask, and the coordinates of a bounding box on the mask are saved for consequent steps. The output of this stage is shown in Figure 2. As the site-information is saved it is possible at any time to invert the processing back to the original image format.

4.2. Crane segmentation

Crane detection is done via interactive segmentation. As a preprocessing step, we smooth the image, filtering it with a Gaussian filter. The user can choose to apply the filter again and again, which equates to filtering with a wider Gaussian kernel (convolving n times a Gaussian with itself increases the resulting standard deviation by a $\sqrt{n}$ factor). This allows the user to remove details and control how fine-grained the segmentation should be.

Cranes produce distinctive bright reflections in the SAR images. However, segmenting them with a global threshold is not possible because both the cranes and the edges of the heaps present high intensity values. This is why only local intensity thresholding is proposed after the image was smoothed. To do this, the user draws a bounding box and selects an intensity threshold to be applied inside it. This is complemented by an area filtering that allows the removal of little connected components. See Figure 2.

4.3. Heap segmentation

Preprocessing. Once the cranes were segmented they are masked out, i.e. fused with the region-of-interest mask. As our heap segmentation tool depends on intensity thresholding, smoothing is again convenient. We apply a non-local-means denoiser [15] as smoother so as to preserve image edges.

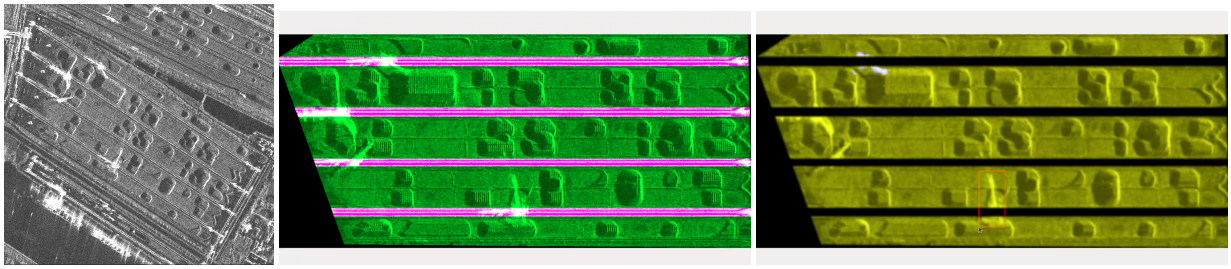

Fig. 2: Left: Example image site. Middle: Site-information processing, rail segmentation step. Right: Crane segmentation using bounding box: top-left of the image shows a segmented crane and bottom-center shows a bounding box to be thresholded.

Algorithm 1 Segment heaps

```
mask_in = merge_masks(site_mask, cranes_mask)
mean_color = i_background_color(img)
img_nl = i_uniformize(img * mask_in, 'nl-means')
img_hwr = abs(img_nl - mean_color) * mask_in
seg = i_segment(img_nl, img)
seg = fill_horizontal(seg)
seg = i_horizontal_merge(img, seg)
seg = i_segment_bbox(img_nl, img, seg)
seg = i_area_filter(img, seg)
final_mask = site_process_inv(seg)
```

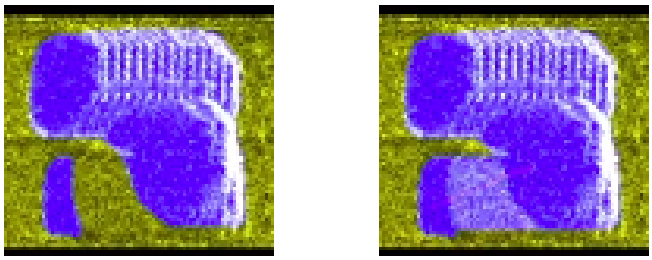

Fig. 3: Horizontal filling before and after merging. A line was drawn between the disconnected components. After the second click, the intermediate area was automatically filled.

The main complication brought by these images to classical segmentation algorithms is that stockpiles are defined by both dark and bright regions, not always connected. We tackle this by half-wave-rectifying the uniformized image. More specifically, the background color is extracted as the mean of a user-defined area (drawing a polygon). Then the absolute difference between the uniformized image and the background color is computed.

Segmentation: The interactive segmentation follows, with a global thresholding step and a local thresholding step for crane segmentation. Both steps are followed by area filtering. The interactive segmentation is done twice: before and after the merging and filling. This allows for more flexibility. After the primary segmentation has been conducted, some segments corresponding to the same stockpile might be separated. Indeed, sometimes the top of the stockpiles has background intensity, and it is hard to differentiate from the background itself without further contextual information. To solve this, a horizontal filling is applied in conjunction with the interactive merging of regions. The horizontal filling runs for each segmented component individually. For a given segmented component, each row of pixels is painted between the first and the last segmented pixels on the row. The process involves interaction by merging. Merging means that the user can draw a line between unconnected components to connect them. After each connection, the horizontal filling takes place and allows to easily segment all pixels between two segmented components (see Figure 3).

The baseline method: The previous segmentation method used to find cranes and stockpiles consisted of a paintbrush with varying sizes and an eraser. This is as flexible as one can get but takes non-negligible time per stockpile or crane. In aggregate, using global operations at least doubles the speed of the method, while eliminating trembling hand blur. The paintbrush approach is nevertheless compatible with the interactive approach here proposed.

Shape from shading: For the SfS step, we use a variational approach that builds upon a linearization of the reflectance map [4]. Because of the global linearization, this method is not very accurate for synthetic examples. Compared to more precise methods, it has the advantage of being very robust to the perturbations occurring in real images.

Our SfS approach expects the observed image to be nadir, so we solve the equation in the range-azimuth plane. In this system of coordinates, the vector $\mathbf{d}$ in (2) is $\mathbf{d} = (1, 0, 0)$. The ground is modeled as a slope whose angle depends on the incidence angle of the satellite. This slope is used to encode Dirichlet boundary conditions. To deal with occluding objects like the cranes we enforce homogeneous Neumann conditions and inpaint the occlusion area with the heat equation. Finally, we project the solution in the UTM grid.

5. RESULTS AND CONCLUSION

Interactive segmentation times: We measured the segmentation time with both the paint-brush tool and the interactive segmentation tool. For the first, the average segmentation time of both cranes and stockpiles was 25 minutes for the reference image. For the second, the average time over 6 images is 11.75 minutes, with extremes being 9.76 min and 13.46 min.

These times can still be reduced. The demo video illustrates the full process, including the first region of interest selection, which is done only once per site. It takes less than 9 minutes.

Masks influence on ground height estimation: The difference between providing masks or using site information only is depicted in Figure 4. The main advantage is that we do not have any incorrect estimation of the ground level, which is forced to be zero in the case using masks. The influence of this is quite drastic: for the reference image, the volume estimated from Figure 4-middle which uses the masks is 2.0, while the volume assigned to the ground in Figure 4-left is -1.4, a negative bias of 70%. When computing the bias

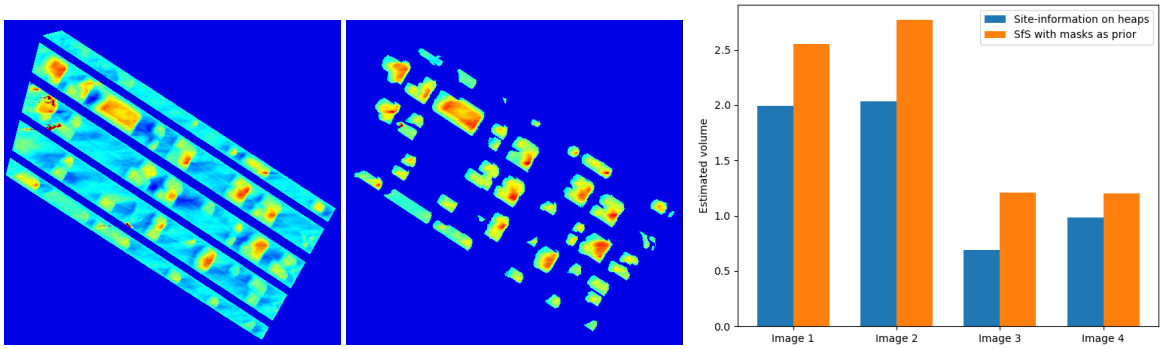

Fig. 4: Height maps yielded by the SfS method using only site information (left), using provided masks (middle), and the comparison between these two cases as total estimated volume over masks per image (right).

introduced by the height estimation of the ground on other images, a similar behavior was found. The minimum absolute bias introduced is 60% while the maximum is 100%. The average bias is -74.6%.

Influence of masks on heaps height estimation: We also note a non-negligible difference between the volume estimate with and without masks, as shown in Figure 4-right. Note that volume estimation is simply a sum, which represents volume up to a constant factor. This constant factor is obtained by calibrating the data with on-site volume measures. The phenomenon observed in Figure 4-right is explained by the effect of ground height estimation. As observed, the ground height estimation introduces a negative bias as the average ground height is negative. Thus, the heaps start from under zero height.

Limitations of the method: This method is useful for an homogeneous material, i.e. with constant texture and reflectance properties. Layover is an issue not addressed here, inherent to SfS methods applied to SAR: it affects slope estimation but not height estimation after the layovered part. We can expect a perturbation in the reconstruction that is relatively minor.

6. CONCLUSION

This paper explored the impact of boundary conditions on SfS applied to SAR images to estimate stockpile volumes. We showed that the rough volume estimation was biased because the volume assigned to the ground is far from zero. At the same time, this bias can be avoided by segmenting the stockpiles and using the masks for determining mixed boundary conditions. We finally proposed an interactive segmentation tool that halves the annotation time.

7. REFERENCES

[1] H. Fawzy, "The accuracy of determining the volumes using close range photogrammetry," *IOSR-JMCE*, vol. 12, pp. 10–15, 2015.

[2] H. He, T. Chen, H. Zeng, and S. Huang, "Ground control point-free unmanned aerial vehicle-based photogrammetry for volume estimation of stockpiles carried on barges," *Sensors*, vol. 19, no. 16, pp. 3534, 2019.

[3] G. Tucci, A. Gebbia, A. Conti, L. Fiorini, and C. Lubello, "Monitoring and computation of the volumes of stockpiles of bulk material by means of uav photogrammetric surveying," *Remote Sensing*, vol. 11, no. 12, pp. 1471, 2019.

[4] M. d'Autume, A. Perry, J.-M. Morel, E. Meinhardt-Llopis, and G. Facciolo, "Stockpile monitoring using linear shape-from-shading on planetscope imagery," *ISPRS Annals*, vol. 2, pp. 427–434, 2020.

[5] B. K. P. Horn, "Shape from shading: A method for obtaining the shape of a smooth opaque object from one view," Tech. Rep. TR-232, Artificial Intelligence Laboratory, MIT, 1970.

[6] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. B. Girshick, "Masked autoencoders are scalable vision learners," *CoRR*, vol. abs/2111.06377, 2021.

[7] M. Chen, T. Artières, and L. Denoyer, "Unsupervised object segmentation by redrawing," *arXiv preprint arXiv:1905.13539*, 2019.

[8] W. Kim, A. Kanezaki, and M. Tanaka, "Unsupervised learning of image segmentation based on differentiable feature clustering," *TIP*, vol. 29, pp. 8055–8068, 2020.

[9] B. Manjunath and R. Chellappa, ," *TPAMI*, vol. 13, no. 5, pp. 478–482, 1991.

[10] B. Settles, "Active learning literature survey," Tech. Rep. TR-1648, Department of Computer Sciences, University of Wisconsin– Madison, 2009.

[11] R. Benenson, S. Popov, and V. Ferrari, "Large-scale interactive object segmentation with human annotators," in *CVPR*, June 2019.

[12] K. Sofiiuk, I. A. Petrov, and A. Konushin, "Reviving iterative training with mask guidance for interactive segmentation," *arXiv preprint arXiv:2102.06583*, 2021.

[13] T. Kontogianni, M. Gygli, J. R. R. Uijlings, and V. Ferrari, "Continuous adaptation for interactive object segmentation by learning from corrections," in *ECCV*, 2020.

[14] J.-D. Durou, M. Falcone, and M. Sagona, "Numerical methods for shape-from-shading: A new survey with benchmarks," *CVIU*, vol. 109, no. 1, pp. 22–43, 2008.

[15] A. Buades, B. Coll, and J.-M. Morel, "Non-local means denoising," *Image Processing On Line*, vol. 1, pp. 208–212, 2011.