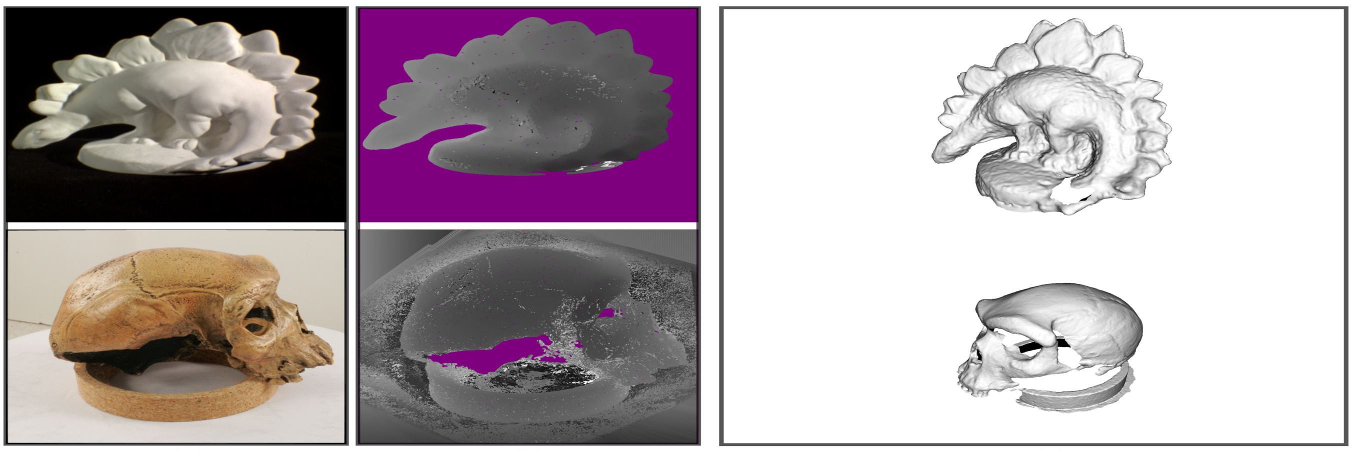
Towards a Stochastic Depth maps Estimation for Textureless and Quite Specular Surfaces

Abdelhak Saouli
Department of Computer Science , LESIA, University of Mohamed Khider
Bikra, Algeria
saouli-abdelhak@hotmail.com

Mohamed Chaouki Babahenini
Department of Computer Science , LESIA, University of Mohamed Khider
Bikra, Algeria
chaouki.babahenini@gmail.com



(a) (b) (c)

Figure 1: (a) A sample of the input images from the MiddleBurry dataset and the nskulla dataset, (b) The resulted depth maps computed using our approach, (c) The final 3D model reconstructed by merging all the depth maps

ABSTRACT

The human brain is constantly solving enormous and challenging optimization problems in vision. Due to the formidable meta-heuristics engine our brain equipped with, in addition to the widespread associative inputs from all other senses that act as the perfect initial guesses for a heuristic algorithm, the produced solutions are guaranteed to be optimal. By the same token, we address the problem of computing the depth and normal maps of a given scene under a natural but unknown illumination utilizing particle swarm optimization (PSO) to maximize a sophisticated photo-consistency function. For each output pixel, the swarm is initialized with good guesses starting with SIFT features as well as the optimal solution (depth, normal) found previously during the optimization. This leads to significantly better accuracy and robustness to textureless or quite specular surfaces.

CCS CONCEPTS

• **Computing methodologies → Reconstruction**; Matching;

SIGGRAPH '18 Posters, August 12-16, 2018, Vancouver, BC,Canada

ACM ISBN 978-1-4503-5817-0/18/08.
https://doi.org/10.1145/3230744.3230762

KEYWORDS

Depth-map, Multi-view Stereo, 3D reconstruction, Textureless, Specular, Swarm Optimization

1 INTRODUCTION

According to the classification of [Seitz et al. 2006], dense multi-view stereo (MVS) reconstruction methods can be categorized in term of how the scene is represented, for example, one popular choice is a volume where the object can be represented via gird of voxels. Another representation is depth maps, the scene is described via the distance between the surface and the center of the projection, methods that use such representation are referred to as multi-view depth map estimation (MVDE) by [Zheng et al. 2014]. MVDE methods can be further categorized into two classes. First, the patch-based reconstruction methods with their local approach to refining the position and orientation of each patch [Furukawa and Ponce 2010] independently which make it extremely powerful in recovering thin details of an object's surface. However textureless, occluded or non-Lambertian surfaces in general still pose a stumbling block for these approaches to achieve completeness and accuracy. On the other hand, variational methods [Langguth et al. 2016] optimize over all the object surface by adding some form of regularization (smoothness term) to deal with textureless areas that are not well matched by classical photo-consistency measurements. However one has to be careful when designing the energy

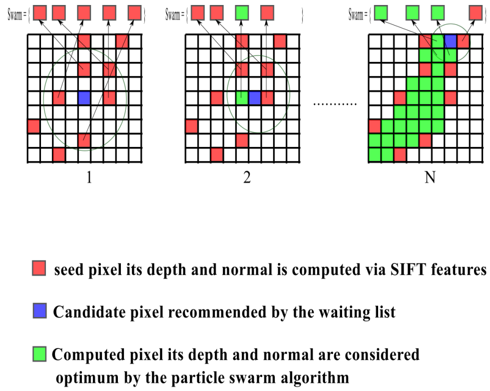


Figure 2: Our algorithm in a nutshell: We compute the depth and normal of each candidate pixel (Blue) using the PSO algorithm, the size of the swarm is dynamic and changes according to the number of optimum pixels (Green)

function to be minimized since solving such a problem requires a derivative based approach such as conjugate gradient method which proved to be hard or even impossible. We show that using the particle swarm optimization our approach is derivative-free and it can recover accurate geometry from textureless areas without any explicit regularization in addition to fine-grained detail of an object's surface.

2 TECHNICAL APPROACH

Consider a scene captured by **M** photographs. In this work, the problem of surface reconstruction is posed as search problem for optimal depth and orientation on one of these images. Similar to every patch-based approach, we set our geometric model to be a *nxn* pixel window that represents a small planar patch. Equation (1) shows the model we followed.

$$\mathbf{X}(u_{ref}) = \mathbf{C}_{ref} + \vec{\mathbf{r}}(u_{ref}) * [\frac{\vec{\mathbf{n}} \cdot \vec{\mathbf{c}}_{ref} + d}{\vec{\mathbf{n}} \cdot \vec{\mathbf{r}}(u_{ref})}] \tag{1}$$

Where $\mathbf{X}(u_{ref})$ represents the 3D position of a given u_{ref} pixel from the refrence view. The center of camera projection, the normalized ray direction through a given pixel and the position vector of the camera center are represented by $\mathbf{C}_{ref}$, $\vec{\mathbf{r}}(u_{ref})$, $\vec{\mathbf{c}}_{ref}$ respectively. Finally the only unknowns in our systems are first,$\vec{\mathbf{n}}$ which is the patch normal vector and d the distance from the origin. The similarity measurement that we adopted as an energy function to be maximized is a combination of a robust Normalized Cross-Correlation (NCC) and the Sum of Absolute Differences (SAD). Such photo-consistency measurement proved to be robust to noises occlusion and other problems.

Starting the matching algorithm, we adopt the constriction factor version of the original particle swarm optimization algorithm to build a dense depth map, note that the dimension of our problem is relatively small and contains only four elements, the (x,y,z) of the normal vector and the distance from the origin d. First, we create a waiting list L that contains pixels we want to compute their relative depth and normal, if the candidate is optimum, we add its four neighbours to L as new possible candidates. Next, We assigned a swarm S to each candidate element from the list, the size of the swarm is related to the radius of the search circle centred around the candidate pixel:

$$radius = \frac{max\_radius}{ln(number\_of\_optimum\_pixels + 1)} \tag{2}$$

The key idea is to construct a 2D selection pool of particles where the position of each particle in the grid corresponds to a pixel in the reference view. Therefore, we choose the closest particles to the candidate pixel inside a search space of a radius computed with equation (2) that will be participating in the optimization process. We fill the pool gradually starting with particles initialized by SIFT features. Then, if a candidate pixel is a successful match, we add a new particle to the pool, The particle vector is filled with the information extracted from this optimum pixel (Figure 2). The process stops when there are no more pixels in the list L

3 THE KEY FINDINGS

For our experiment we used multiple datasets, First the dino dataset it contains images of dino plaster captured under the lab condition, the object has a textureless Lambertian surface, therefore, the matching process is extremely difficult. The skull datasets are specular and the lighting conditions changes in each view. Such datasets pose a problem to the MVS algorithms and produce incomplete depth maps. We claim that our approach can deal with such problem as shown in (Figure 1).

Our first key finding is that the use of the swarm optimization smartly mimics the variational approaches; however, it is less complicated, more efficient and derivative free. Our goal is to avoid explicit regularization. Instead, the use of the swarm formulation and the neighbouring strategy to initialize each particle, in addition to the use of the waiting list to propagate in the image space guaranteed the smoothness of the surface implicitly while preserving thin details.

Currently, our depth map estimation is based solely on a stochastic stereo matching algorithm. We should be able to refine and denoise these depth maps while taking into account the orientation map obtained. We intend to compare our method with the state-of-art approaches on the Middlebury website, We hypothesize that our method will result in a high completeness score and comparable accuracy to those of the current top performing algorithms. also, we are working on a faster implementation and an optimized code.

REFERENCES

Yasutaka Furukawa and Jean Ponce. 2010. Accurate, Dense, and Robust Multiview Stereopsis. *IEEE Trans. Pattern Anal. Mach. Intell.* 32, 8 (2010), 1362–1376. https://doi.org/10.1109/TPAMI.2009.161

Fabian Langguth, Kalyan Sunkavalli, Sunil Hadap, and Michael Goesele. 2016. Shading-Aware Multi-view Stereo. In *Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III.* 469–485. https://doi.org/10.1007/978-3-319-46487-9_29

Steven M. Seitz, Brian Curless, James Diebel, Daniel Scharstein, and Richard Szeliski. 2006. A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2006), 17-22 June 2006, New York, NY, USA.* 519–528. https://doi.org/10.1109/CVPR.2006.19

Enliang Zheng, Enrique Dunn, Vladimir Jojic, and Jan-Michael Frahm. 2014. PatchMatch Based Joint View Selection and Depthmap Estimation. In *2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, OH, USA, June 23-28, 2014.* 1510–1517. https://doi.org/10.1109/CVPR.2014.196