MARTIN-LUTHER-UNIVERSITÄT
HALLE-WITTENBERG

3D Data Derivatives of Grotta di Fumane: GigaMesh-processed, Annotations and Segmentations

Supplementary Documentation

Florian Linsel, Jan Philipp Bullenkamp and Hubert Mara

September 1, 2023

Martin-Luther University of Halle-Wittenberg
Institute of Computer Science
Halle, 06114, Germany

Abstract

This documentation will serve as a description of our dataset, which contains derivatives of the open-access publication "The Open Aurignacian Project. Volume 1: Fumane Cave in northeastern Italy" by Falcucci and Peresani [FP22]. Our derived dataset is used to showcase our segmentation algorithm [BHM23], shown in "Lithic Feature Identification in 3D based on Discrete Morse Theory" [BLM22], in "Discrete Morse Theory Segmentation on High-Resolution 3D Lithic Artifacts" [BLM23], in "Linking Scars: Topology-based Scar Detection and Graph Modeling of Paleolithic Artifacts in 3D" [LBM23] and will serve as a benchmark dataset for future analyses. Up to this moment and as far as our knowledge goes, our dataset is the first dataset of annotated lithic artifacts. For all basic archaeological information concerning the archaeological interpretation of the artifacts, we want to refer to the original data publication [FP22].

Contents

1 Methods

Our applied methods consists of three main procedures. First, the preprocessing in GigaMesh, including calculating the MSII curvature values, second, a manual annotation of the 3D meshes and third, the results of our Morse theory based segmentation algorithm.

1.1 Preprocessing Pipeline in GigaMesh

In a preprocessing routine, following [MB19], all 732 3D models got manually orientated, cleaned and filled ("_GMOCF") [Mar+10]. Afterwards, we used the Multi-Scale Integral Invariants filter (MSII) to approximate the surface curvature of all artifacts (_r1.00_n4_v256). The MSII filter calculates curvature measures at various scales, namely sphere radii, resulting in a feature vector for each vertex. GigaMesh offers various functions for computing a scalar value for each vector, such as the Euclidean length. In the subsequent sections, the MSII renderings used the maximum function [MK13].

1.2 Annotation

62 artifacts were selected to cover the varying quality of the 3D meshes and the scar-ridge complexity. 57 artifacts come from the A1-A2 unit, which is dated between 42,000 and 39,250 cal BP. To gain a broader perspective, we included also 5 artifacts from the D3 unit (D3b, D3d and D3d base), which is dated between 40,000 and 37,750 cal BP [Hig+09; Mar+23].

We annotated these artifacts manually using Meshlab's paintbrush tool [Cig+08]. Each color represents a label, which we translated to numbered labels (connected components). We used these labels as ground truth dataset ("_gt_labels.ply", "_gt_labels.txt") to evaluate the segmentation results.

1.3 Segmentation Results

We used our segmentation pipeline [BLM22; BLM23] to segment those 62 artifacts that have manual annotations to evaluate our method. Each segmentation requires four parameters: persistence ("_0.06P"), high and low ridge detection thresholds ("_0.07H_0.06L") and a merging threshold ("_0.2M"). We ran the pipeline for a range of parameters and include only the result of the best parameter combination for each artifact.

2 Data Records and Structure

We separated the dataset in 4 zip files, namely the with GigaMesh preprocessed datasets (n = 732), the annotated (n = 62), the segmentated (n = 62) and an example dataset for GdF 31, which resembles the structure of the 3 zip files mentioned above.

2.1 GigaMesh

Within the "GigaMesh.zip" file, there are several subfolders named after the containing ID ranges such as "15-1534", "1537-4114" or "4124-5301", which are included in the folder. Each of these subfolders contains three additional subfolders: "gigamesh-info", "plys" and "views". The gigamesh-info contains ".info.txt" files, which describes the properties of each mesh. In views there are two subfolders, figs, which contains the 6 rendered views as "png" files, and pdfs, which contains a data description, including the 6 images.

2.2 Annotation

In "Annotation.zip" are two subfolders: "plys" and "txts", containing the labeled meshes and the labels as txt files. The subfolder "views" contains all images of the ground truth labels and all pdf files of the reports.

```
Annotation.zip
├── plys
├── txts
└── views
    ├── figs
    └── pdfs
```

Table 1. Folder structure of Annotation.zip

2.3 Segmentation

Under the "Segmentation.zip" folder, we limit the scope of the data to one subfolder "txts". This subfolder contains the segmentation results as txt files.

```
Segmentation.zip
└── txts
```

Table 2. Folder structure of Segmentation.zip

3 Conventions

3.1 Name Conventions

Due to the intrinsic complexity of the dataset, consisting of 2D renderings as well as 3D meshes, we want to include a chapter introducing a naming convention, which is based upon the GigaMesh naming convention [Mar+10; Mar19]. We are keeping the original ID coming from [FP22] as name and add standardized endings, referencing specific processing steps.

3.1.1 GigaMesh.zip

In the first subchapter, we will briefly describe the GigaMesh naming convention presented in Mara, Krömker, Jakob, and Breuckmann [Mar+10] (Ch. 5.1). To give an idea, here are the main conventions:

"_GMOCF" is the main preprocessing routine in GigaMesh:

- "_GMOCF" e.g. <filename>**_GMOCF**_r1.00_n4_v256.ply
  - "_GM" = GigaMesh Software Framework
  - "O" = manually orientated
  - "C" = cleaned
  - "F" = filled

"_r1.00_n4_v256" involves the MSII filter algorithm [MK13]:

- "_r1.00_ n4_v256" e.g. <filename>_GMOCF**_r1.00_n4_v256**.ply
  - "_r1.00" = maximum radius of the spheres
  - "_n4" = 2^4 = 16 equidistant radii scales in the interval [0, r]
  - "_v256" = a rasterization of the volume with, e.g., means that there are 256 voxels along the diameter of the biggest sphere

3.1.2 Annotation.zip

This zip folder consists of all annotated artifacts:

- "_gt_labels" e.g. <filename>_GMOCF**_gt_labels**.ply
  - "_gt_labels" = ground truth labels

3.1.3 Segmentation.zip

The results of our segmentation method are stored in this zip folder, containing ".txt" files with the parameters written into the filenames.

- "_0.06P_0.07H_0.06L_0.2M" e.g. <filename>**_0.06P_0.07H_0.06L_0.2M**.txt
  - "_0.06P" = 0.06 persistence
  - "_0.07H" = 0.07 high ridge detection threshold
  - "_0.06L" = 0.06 low ridge detection threshold
  - "_0.2M" = 0.2 merging threshold

3.2 Name Convention for Reports

The pdf reports directly refers to the artifact with the first part of its name. The attached extension refers to the type of mapped property on the surface of the artifact.

- "_MSII" e.g. <filename>_GMOCF_r1.00_n4_v256.volume**_MSII**.pdf
  - "_MSII" = refers to the property, which is represented as color variable on the surface of the mesh.

3.3 Name Convention for Images

All images are named according to their reference to the report.

- "_01_va_top" e.g. <filename>_GMOCF_r1.00_n4_v256.volume_MSII **_01_va_top**.png
  - "_01" = number of the image according to the GigaMesh image order for creating the GigaMesh report.
  - "_va" = mesh was rotated around the vertical axis.
  - "_top" = refers to the direction of the rendering relativ to the 3D mesh, also known in this example as top view.

3.4 Views Convention

All renderings are created using the new GigaMesh tool "Directory: Render from all files". This tool allows the user to select a directory containing ply files. It then creates an overview report from all the ply files in the directory, using a Latex template.

All 3D meshes get rendered using the first angle orthographic projection (DIN ISO 128-30-1)[SMP20], which is a depiction method of representing a three-dimensional object using six views. These views consists of the front, top, right-side, left-side, bottom, and back views.

Additionally to the views, two tables with the principle parameters are included in the report. The left table includes the maximum dimension of the bounding box (Max. dim. width x height x depth), the number of vertices (Vertices) and faces (Faces). The right table consists of the surface acquisition (Surface, acq.), the average resolution (Res., avg.) in cm^{-2} and DPI, as well as the estimated volume (Volume, est.). In the bottom right of each page is the scale factor included, referencing whether the mesh needed to be rescaled.

All reports are provided with a Creative Commons "Attribution-ShareAlike 4.0 International" license attached as second page.

4 GdF 31 (example_31.zip)

For a better understanding of the data structure, we use GdF 31 as an example to show the overall structure of the dataset (Fig 4).

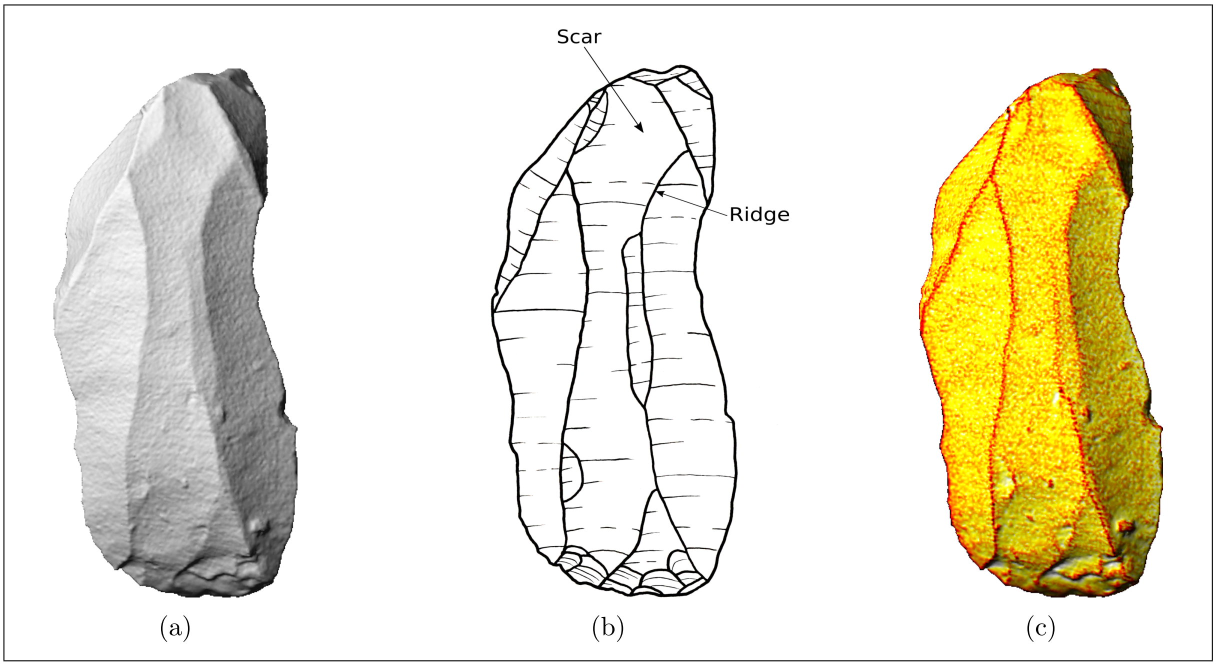

Figure 1. Images of GdF 31, (a) front view of the solid 3D mesh; (b) drawing; (c) MSII-curvature mapped as function value.

4.1 Folder structure

In this chapter, we will describe the structure of our dataset using the artifact GdF 31. The example is provided in an additional example zip file (example_31.zip). This zip file resembles the overall structure of the other three zip files and therefore contains a GigaMesh, an Annotations and a Segmentation folder. To give an idea of the overall data containing folders, we will focus on the overview pdf files provided in the pdfs subfolders.

```
example_31.zip
 ├── GigaMesh
 │    └── ..
 ├── Annotation
 │    └── ..
 └── Segmentation
      └── ..
```

Table 3. Folder structure of example_31.zip

4.2 Subfolder for Renderings with GigaMesh

All renderings in the views folder are based on only one ply file. The differences in the rendered images can be condensed into two properties: the illumination and the surface color. The surface color can be colored by either a continuous function value e.g. maximum MSII or by a label e.g. the ground truth labels. Note that GigaMesh provides only 10 distinct label colors, i.e. if the number of labels exceeds 10, it repeats itself. Most of the renderings are not additionally illuminated. While the MSII views are not illuminated, the solid and the MSII_light images are illuminated by a virtual light source from the upper left corner ($\Theta : 140$; $\Phi : 40$).

4.2.1 Solid Renderings with GigaMesh

Fig 2 shows the illuminated solid view rendering.

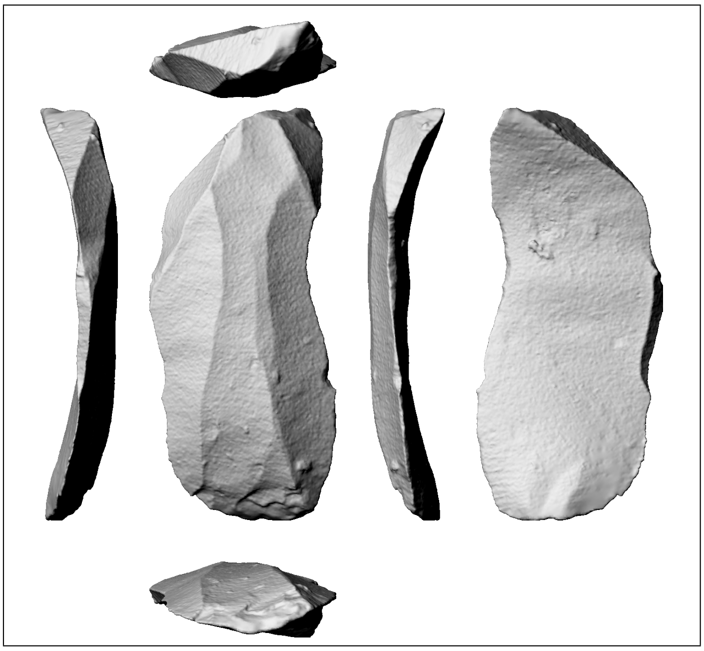

Figure 2. Solid view of the artifact GdF 31

4.2.2 MSII Renderings with GigaMesh

Fig 3 shows the MSII values mapped as function values.

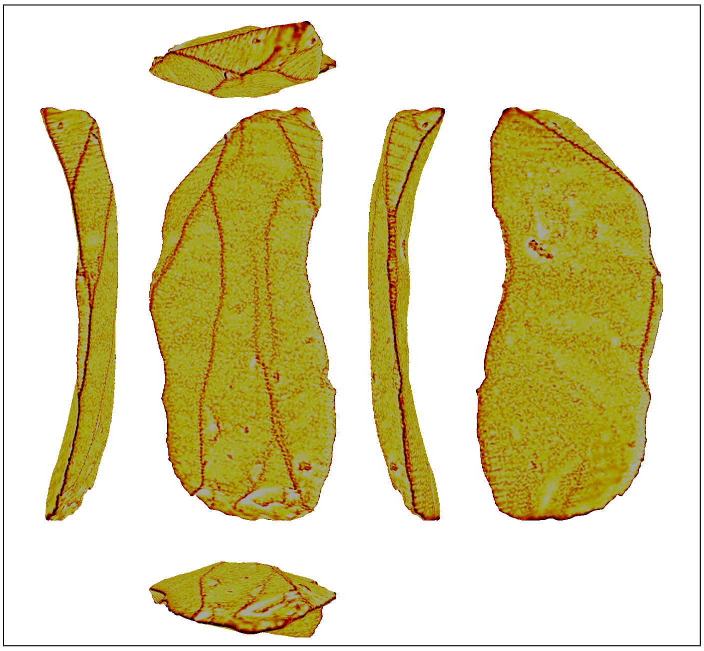

Figure 3. MSII view of the artifact GdF 31

4.2.3 MSII_light Renderings with GigaMesh

Fig 4 shows the illuminated MSII values mapped as function values.

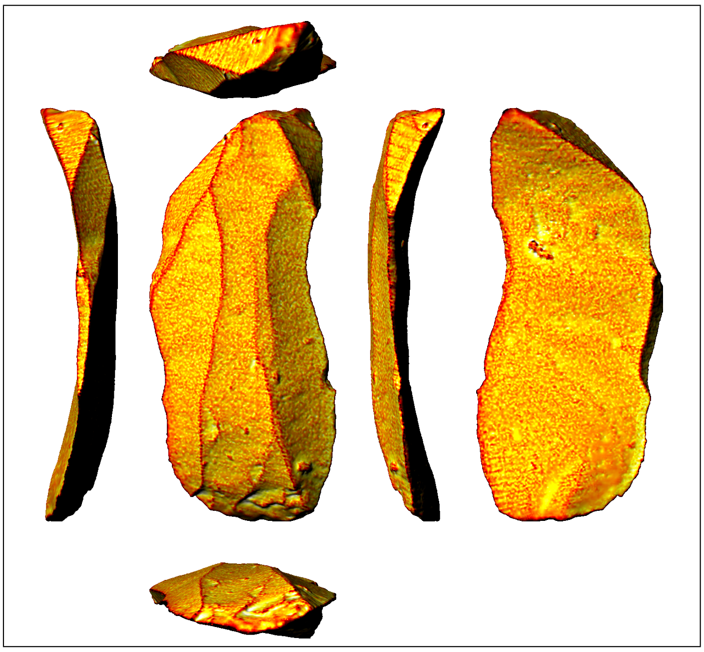

Figure 4. MSII light view of the artifact GdF 31

4.2.4 Annotation Renderings with GigaMesh

Fig 5 shows ground truth labels mapped as labels.

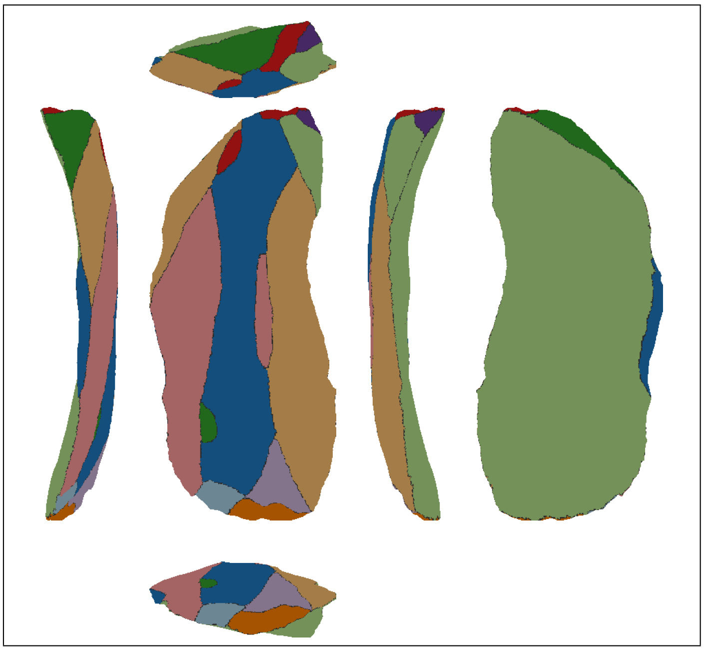

Figure 5. GT label view of the artifact GdF 31

4.2.5 Segmentation Renderings with GigaMesh

Fig 6 shows the result of the Morse Theory segmentation pipeline mapped as labels.

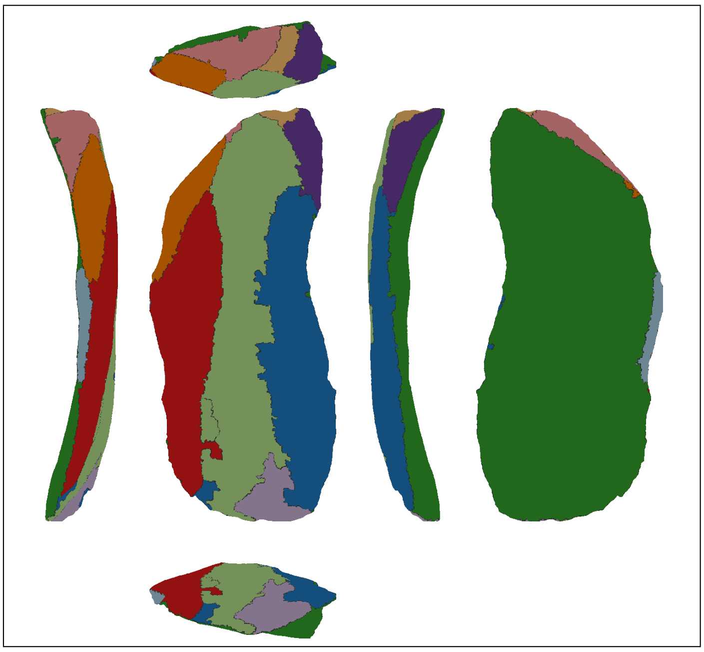

Figure 6. Segmentation view of the artifact GdF 31

4.2.6 Views

All views depicting GdF 31 include, as mentioned in the views convention, 6 images, two tables with basic parameters (Fig 7) and on the second page a Creative Commons “Attribution-ShareAlike 4.0 International” license.

Fig 7 shows for the example of GdF 31, the two parameter tables, the reference to GigaMesh and the scale of the images.

Max. dim.	1,7 × 3,8 × 0,7 cm	Surface, acq.	1 170 mm^2
Vertices Faces	146 174 292 344	Res., avg.	12 500 cm^{-2} 284 DPI
Material	original, lithic	Volume, est.	1 cm^3

Rendered with HTTPS://GIGAMESH.EU Scale 1 : 0.7

Figure 7. Parameter tables of GdF 31

5 Folder Trees

5.1 Folder Tree for GigaMesh.zip

```
GigaMesh.zip
├── gigamesh_info
├── 15-1534
│   ├── plys
│   └── views
│       ├── MSII
│       ├── MSII_light
│       └── solid
├── 1537-4114
│   ├── plys
│   └── views
│       ├── MSII
│       ├── MSII_light
│       └── solid
├── 4124-5301
│   ├── plys
│   └── views
│       ├── MSII
│       ├── MSII_light
│       └── solid
├── 5302-5757
│   ├── plys
│   └── views
│       ├── MSII
│       ├── MSII_light
│       └── solid
├── 5758-5943
│   ├── plys
│   └── views
│       ├── MSII
│       ├── MSII_light
│       └── solid
├── 5945-8268
│   └── ...
├── 8269-10422
│   └── ...
└── 10423-10455
    └── ...
```

See the 5945-8268, 8269-10422 and 10423-10455 tree structure on the next page:

```
GigaMesh.zip
├── 15-1534
│   └── ...
├── 1537-4114
│   └── ...
├── 4124-5301
│   └── ...
├── 5302-5757
│   └── ...
├── 5758-5943
│   └── ..
├── 5945-8268
│   ├── plys
│   └── views
│       ├── MSII
│       ├── MSII_light
│       └── solid
├── 8269-10422
│   ├── plys
│   └── views
│       ├── MSII
│       ├── MSII_light
│       └── solid
└── 10423-10455
    ├── plys
    └── views
        ├── MSII
        ├── MSII_light
        └── solid
```

5.2 Folder Tree for example_31.zip

```
example_31.zip
 └── GigaMesh
      ├── gigamesh_info
      │    └── 31_GMOCF_r1.00_n4_v256.info.txt
      ├── plys
      │         └── 15-1534
      │              └── 31_GMOCF_r1.00_n4_v256.volume.ply
      └── views
           ├── 15-1534
           ├── MSII
           │    ├── figs
           │    │    ├── 31_GMOCF_r1.00_n4_v256.volume_MSII_01_va_top.png
           │    │    ├── 31_...._02_va_left.png
           │    │    ├── 31_...._03_va_front.png
           │    │    ├── 31_...._04_va_right.png
           │    │    ├── 31_...._05_va_bottom.png
           │    │    └── 31_...._06_va_back.png
           │    └── pdfs
           │         └── 31_GMOCF_r1.00_n4_v256.volume_MSII.pdf
           ├── MSII_light
           │    ├── figs
           │    │    ├── 31_GMOCF_r1.00_n4_v256.volume_MSII_light_01_va_top.png
           │    │    ├── 31_GMOCF_...._02_va_left.png
           │    │    ├── 31_GMOCF_...._03_va_front.png
           │    │    ├── 31_GMOCF_...._04_va_right.png
           │    │    ├── 31_GMOCF_...._05_va_bottom.png
           │    │    └── 31_GMOCF_...._06_va_back.png
           │    └── pdfs
           │         └── 31_GMOCF_r1.00_n4_v256.volume_MSII_light.pdf
           └── solid
                ├── figs
                │    ├── 31_GMOCF_r1.00_n4_v256.volume_solid_01_va_top.png
                │    ├── 31_...._02_va_left.png
                │    ├── 31_...._03_va_front.png
                │    ├── 31_...._04_va_right.png
                │    ├── 31_...._05_va_bottom.png
                │    └── 31_...._06_va_back.png
                └── pdfs
                     └── 31_GMOCF_r1.00_n4_v256.volume_solid.pdf
 ├── Annotation
 │    └── ...
 └── Segmentation
      └── ...
```

See the annotation and segmentation tree structure on the next page:

```
example_31.zip
├── GigaMesh
│   └── ...
├── Annotation
│   ├── plys
│   │   └── 31_GMOCF_gt_labels.ply
│   ├── txts
│   │   └── 31_GMOCF_gt_labels.txt
│   └── views
│       ├── figs
│       │   ├── 31_GMOCF_gt_labels_01_va_top.png
│       │   ├── 31_...._02_va_left.png
│       │   ├── 31_...._03_va_front.png
│       │   ├── 31_...._04_va_right.png
│       │   ├── 31_...._05_va_bottom.png
│       │   └── 31_...._06_va_back.png
│       └── pdfs
│           └── 31_GMOCF_gt_labels.pdf
└── Segmentation
    └── txts
        └── 31_0.06P_0.07H_0.06L_0.2M.txt
```


Acknowledgements

We would like to thank Ernst Max Stötzner for his contributions to GigaMesh and the data preparation in GigaMesh.

Competing interests

The authors declare no competing interests.

References

[BHM23] Jan Philipp Bullenkamp, Theresa Häberle, and Hubert Mara. *MorseMesh*. Version v0.1. Aug. 2023. DOI: 10.5281/zenodo.8297368.

[BLM22] Jan Philipp Bullenkamp, Florian Linsel, and Hubert Mara. "Lithic Feature Identification in 3D based on Discrete Morse Theory". In: *Eurographics Workshop on Graphics and Cultural Heritage*. Ed. by Federico Ponchio and Ruggero Pintus. The Eurographics Association, 2022, 055–058 pages. ISBN: 978-3-03868-178-6. DOI: 10.2312/gch.20221224.

[BLM23] Jan Philipp Bullenkamp, Florian Linsel, and Hubert Mara. "Discrete Morse Theory Segmentation on High-Resolution 3D Lithic Artifact". Submitted to it - Information Technology, 2023.

[Cig+08] Paolo Cignoni, Marco Callieri, Massimiliano Corsini, Matteo Dellepiane, Fabio Ganovelli, and Guido Ranzuglia. "MeshLab: an Open-Source Mesh Processing Tool". In: *Eurographics Italian Chapter Conference*. 2008, pp. 129–136. DOI: 10.2312/LocalChapterEvents/ItalChap/ItalianChapConf2008/129-136.

[FP22] Armando Falcucci and Marco Peresani. *The Open Aurignacian Project. Volume 1: Fumane Cave in northeastern Italy*. Version 1.0.0. Zenodo, 2022. DOI: 10.5281/zenodo.6362150.

[Hig+09] Thomas Higham, Fiona Brock, Marco Peresani, Alberto Broglio, Rachel Wood, and Katerina Douka. "Problems with radiocarbon dating the Middle to Upper Palaeolithic transition in Italy". In: *Quaternary Science Reviews* 28.13 (2009), pp. 1257–1267. DOI: 10.1016/j.quascirev.2008.12.018.

[LBM23] Florian Linsel, Jan Philipp Bullenkamp, and Hubert Mara. "Linking Scars: Topology-based Scar Detection and Graph Modeling of Paleolithic Artifacts in 3D". Submitted to PCIArchaeology, 2023. DOI: 10.5281/zenodo.8296270.

[Mar+10] Hubert Mara, Susanne Krömker, Stefan Jakob, and Bernd Breuckmann. "GigaMesh and Gilgamesh 3D Multiscale Integral Invariant Cuneiform Character Extraction". In: *VAST: International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage*. Ed. by Alessandro Artusi, Morwena Joly, Genevieve Lucet, Denis Pitzalis, and Alejandro Ribes. The Eurographics Association, 2010. ISBN: 978-3-905674-29-3. DOI: 10.2312/VAST/VAST10/131-138.

[Mar+23] Ana B Marín-Arroyo, Gabriele Terlato, Marco Vidal-Cordasco, and Marco Peresani. "Subsistence of early anatomically modern humans in Europe as evidenced in the Protoaurignacian occupations of Fumane Cave, Italy". In: *Scientific Reports* 13.1 (2023), p. 3788. DOI: 10.1038/s41598-023-30059-3.

[Mar19] Hubert Mara. *HeiCuBeDa Hilprecht - Heidelberg Cuneiform Benchmark Dataset for the Hilprecht Collection*. Version V2. 2019. DOI: 10.11588/data/IE8CCN.

[MB19] Hubert Mara and Bartosz Bogacz. "Breaking the Code on Broken Tablets: The Learning Challenge for Annotated Cuneiform Script in Normalized 2D and 3D Datasets". In: *2019 International Conference on Document Analysis and Recognition (ICDAR)*. 2019, pp. 148–153. DOI: 10.1109/ICDAR.2019.00032.

[MK13] Hubert Mara and Susanne Krömker. “Vectorization of 3D-Characters by Integral Invariant Filtering of High-Resolution Triangular Meshes”. In: *2013 12th International Conference on Document Analysis and Recognition*. 2013, pp. 62–66. DOI: 10.1109/ICDAR.2013.21.

[SMP20] Colin H. Simmons, Dennis E. Maguire, and Neil Phelps. “Principles of first and third angle orthographic projection”. In: *Manual of Engineering Drawing*. Elsevier, 2020, pp. 49–70. ISBN: 978-0-12-818482-0. DOI: 10.1016/B978-0-12-818482-0.00006-2.