TU Informatics

Visualization-Guided Classification of Carbonized Seeds from Early Human Civilizations

Andreas Gogel
Visual Computing

TU Wien Informatics
Visual Computing and Human-Centered Technology
Research Area Computer Graphics
Supervisor: Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Meister Eduard Gröller
In Coorperation with the Department of Botany and Biodiversity Research of the University of Vienna
Contact: andreas.gogel@gmx.at

Motivation

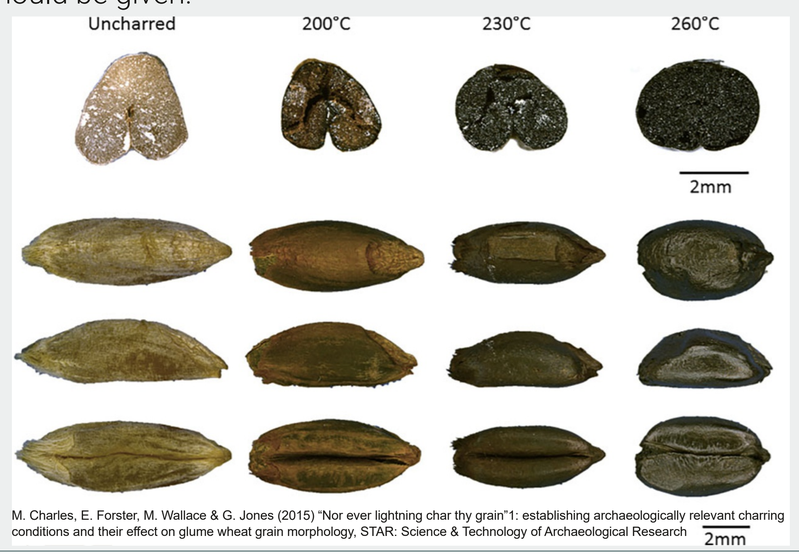
Approximately about 10.000 years ago, humans began the transition from a hunter and gatherer to a settlement society. This process is called the Neolithic Revolution and one of the hot spots of this shifting lies in the Fertile Crescent located in the Near East. Researchers in the field of Botanical Archeology are trying to understand which plant species were selected and how they changed gradually. Most of the remaining seeds from these very early times are carbonized. This way, they could be preserved. However, in this condition, it is hard to identify the species of the seeds. During the burning process, the seeds are deformed and loose surface details. Normally, an expert is consulted to classify them. Since there are only a few experts in this field, an automatic approach is requested.

Tasks

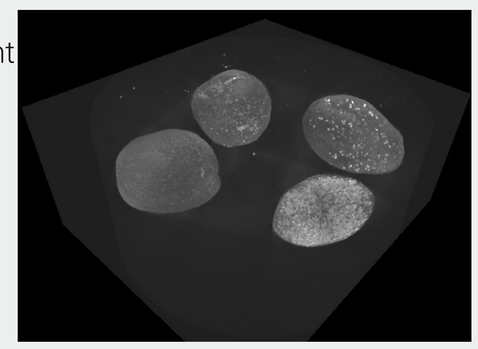
- Create a descriptor of every seed (1)
- Offer visualizations to inspect the CT volumes and the descriptors to find for examlpe outliers and scanning errors (2)
- Offer classification suggestions of seeds from unknown species (with probabilities) (3)
- Offer control mechanisms for the suggestions of the classifier (4)

Data Challenges

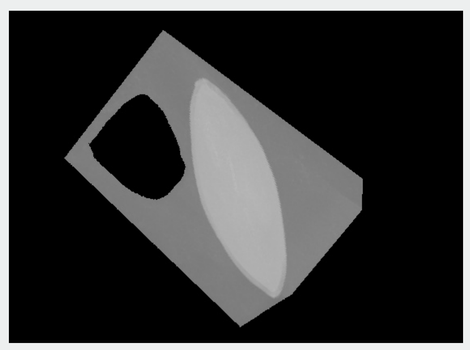
The data set (over 400 GB) consists 1043 labeled seeds of 22 different species (scanning process ongoing).
The distribution of the samples is very unbalanced (some species contains over 200 seeds, others only a single seed). They can be divided into archeological seeds, fresh recent seeds and in the lab carbonized seeds to increase the ground truth. Since some species are already extincted, at least a suggestion of the biological genus should be given.



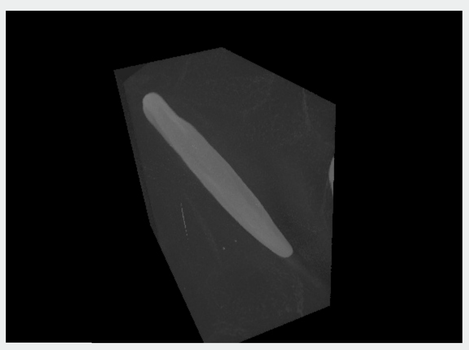
M. Charles, E. Forster, M. Wallace & G. Jones (2015) "Nor ever lightning char thy grain"1: establishing archaeologically relevant charring conditions and their effect on glume wheat grain morphology, STAR: Science & Technology of Archaeological Research



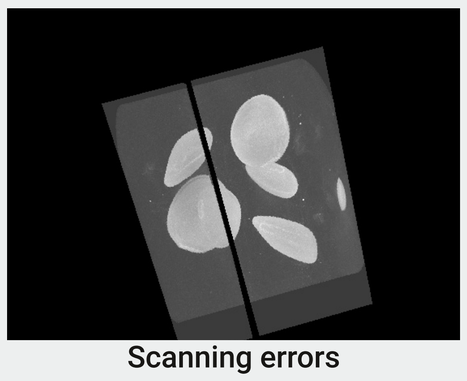
Multiple seeds in one file



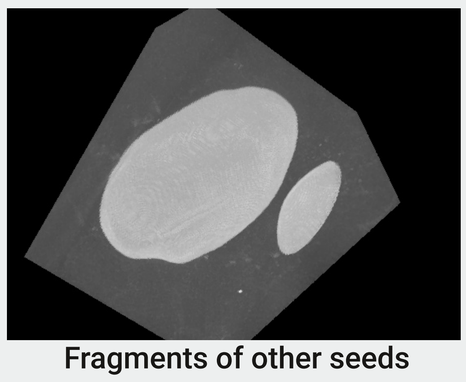
Holes within the samples



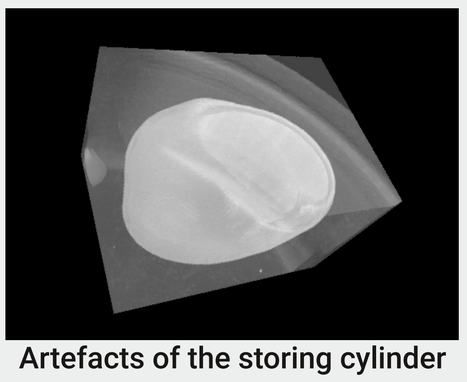
Physically deformed seeds



Scanning errors



Fragments of other seeds

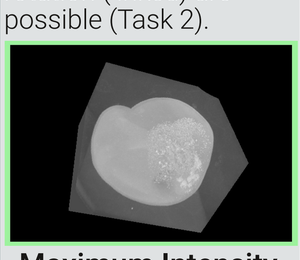


Artefacts of the storing cylinder

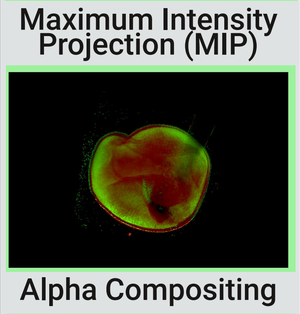
Results

Selection Widget

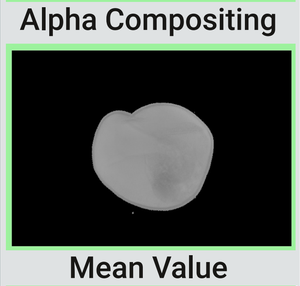
Shows the seed selected in the scatter plot or directly loaded data. Different rendering methods, zooming and rotation (linked) are possible (Task 2).



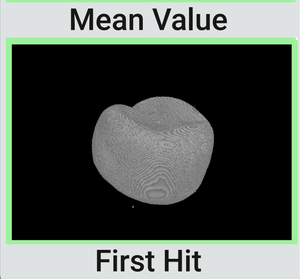
Maximum Intensity Projection (MIP)



Alpha Compositing



Mean Value



First Hit

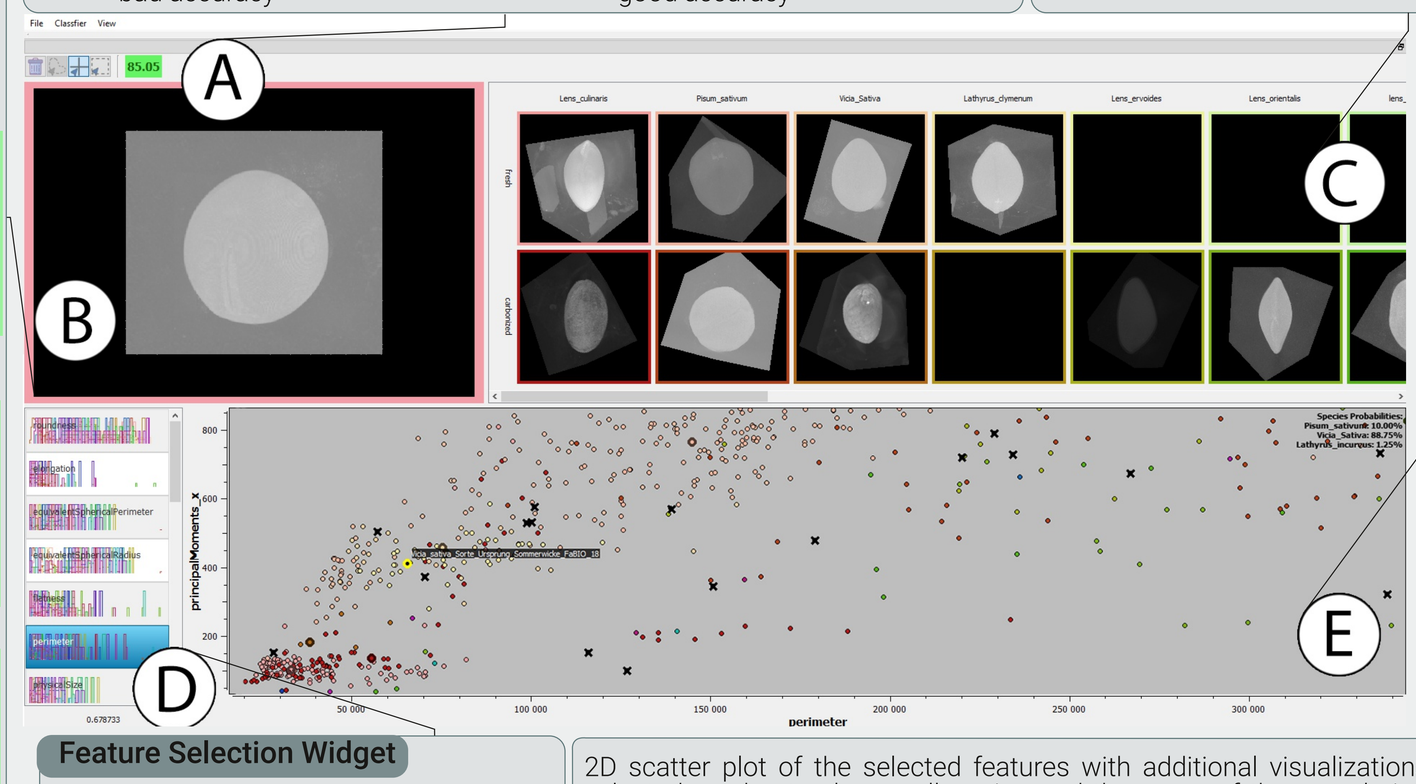
Tool Widget

- Switch between single and multiselection (scatter plot)
- Delete selected samples (scatterplot) e.g., outliers or incorrectly segmented seeds
- Accuracy of the classifier with color coded background

bad accuracy — good accuracy

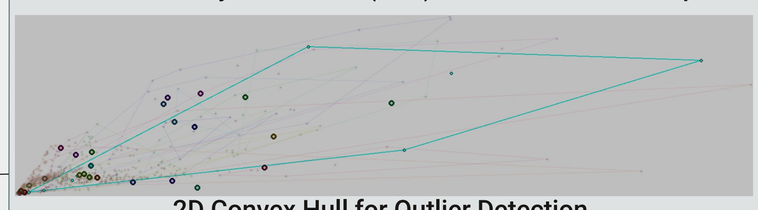
Species Overview Widget

Shows the median seed of every species (fresh and carbonized) according to the selected features with color coded frames. The median is selected to reduce the influence of outliers. The seeds are all equally rotated. Rotation (linked between all widgets) and zooming is possible. If there are no fresh or carbonized samples available, the widget stays black. Selections of species affect also the scatter plot. Different rendering methods are possible (Task 2).

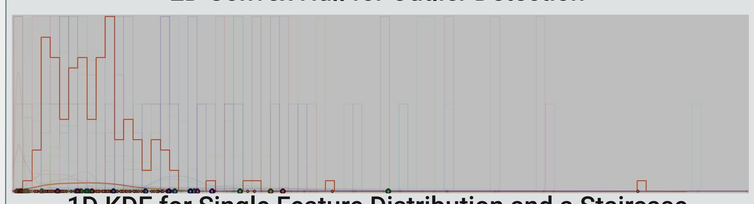


Scatter Plot Widget

2D Kernel Density Estimation (KDE) for Distribution Hot Spots



2D Convex Hull for Outlier Detection



1D KDE for Single Feature Distribution and a Staircase Histogram for Outlier Detection

Feature Selection Widget

Shows a list of all calculated 22 shape features (Task 1) for axis selection. Shape features are for example the elongation, roundness and the volume of the seed. Every button contains also a distribution of the samples of the respective feature. If a new axis is selected, the scatter plot and the median volumes in the species overview widget change.

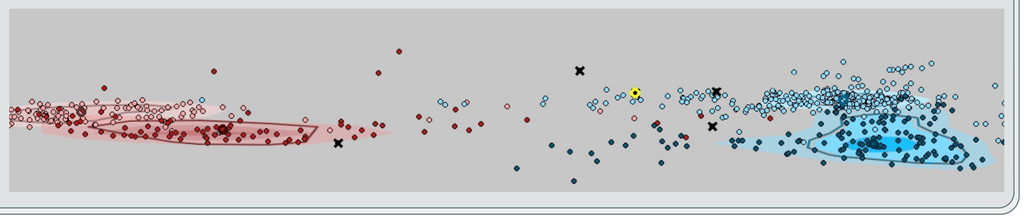
2D scatter plot of the selected features with additional visualizations (Task 2). Median seeds are marked with a colored ring. The selected seed gets also a yellow ring and the name of the sample is displayed. During the cross-validation step incorrectly classified samples are visualized with a black 'x' to show where they are located (Task 4).

Suggestions of the classifier (soft voting, containing a decision tree and a random forest) are shown in the right upper box with the probabilities greater than zero of all classes (Task 3).

Example classifications

Pisum_sativum: 10.00% Vicia_Sativa: 88.75% Lathyrus_incuvus: 1.25%	Species classified
Vicia_Sativa: 12.15% Lathyrus_incuvus: 1.25% Lens_ervoides: 40.73% Lens_orientalis: 45.87%	At least biological genus classified (Lens: 86.6%)

To control and manually improve the result of the classifier, the surrounding seeds can be inspected in the selection widget or the Linear Discriminant Analysis (LDA) view can be used. The LDA view is only calculated for the selected species of the overview widget. This way, this visualization can be improved by selection or deselection of species. In the Image below, the 'x' samples can be classified to the right class by hand (Task 4).



Conclusion

The features, the trained model and separated seeds can be exported (Task 1) and the classifier has an accouracy of approximatly 85% (90% on genus level). The domain expert can check and manualy correct the suggestions of the classifier and is in used for an international project.