Towards Computer Assisted Petrographic Analysis

[1]Alec Iacobucci, Mikhail Echavarri
[1]Department of Anthropology

DigAR Lab

Background

Analysis of ceramic tempers and pastes remains a qualitative method for pottery analysis. This means that results from these analyses are dependent the researcher's experience and that datasets are difficult to compare [1] . This project leverages python image analysis methods to generate rapid quantitative data. We argue that these outputs can provide archaeologists with large scale and repeatable datasets that aid intuitive inferences of ceramic assemblages. We do not intend for these methods to replace traditional and expert analyses, as these require an in depth understanding of a variety of materials.

Figure 1. Ceramic Vessels vs Ceramic Sherds

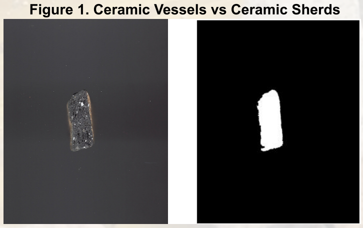


Figure 2. Raw Sherd Scan after cutting and its resulting mask

Aim 1: To create a set of robust, internally consistent functions that collect multivariate data from ceramic sherd scans including: inclusion size, inclusion count, inclusion size distribution, dominant colors, sherd area, inclusions / cm^2

Aim 2: To use multivariate analysis methods to derive pottery groupings from collected data

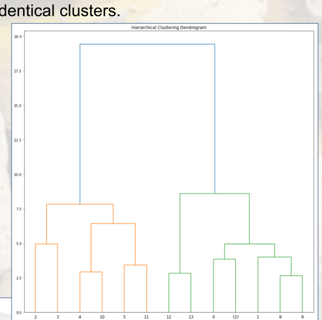
Aim 3: To use these groupings to refine ceramic classifications or create new classifications for currently broad categories

Methods

Once sherds are collected they are measured, weighed, cut in half using a dremel and scanned using a flatbed scanner. Care should be used to make sure sherds are cut along the same plane and to reduce any overhang [2]. These scans are kept at a uniform 600dpi 3316x2680. Using SciKit image, the sherds are masked to remove any background dust that may be recognized as inclusions. Four separate blob detection algorithms are run: Laplacian of Gaussian, Difference of Gaussian, Determinant of Hessian, and a mask based region property function. These are also included in the SciKit Image package. An external package, ColorThief, is used to derive RGB values for the dominant colors of each sherd scan. These are merged with the blob values. The resulting DataFrame is cleaned of redundant data and a PCA function is run. The PCA takes the large variable space and reduces them into lower dimensional space [3]. We then test the visually assessed groups using mahalanobis distance measures, a common distance measure in multivariate space [4]. Hierarchical clustering is also performed.

Results

Though 14 sherds is a small dataset, the results are surprisingly consistent. Previous hand sorting yielded two or four clusters based on overall appearance, and level of discrimination. Multivariate analysis on the data produced via blob detection not only gave the same number of clusters, but organized the sherds into nearly identical clusters.



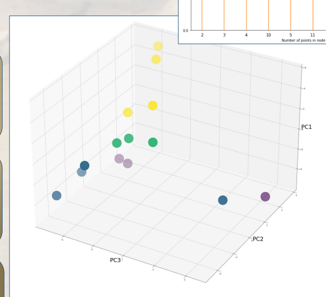


Figure 3: HC Dendrogram. Principal components of sherd data in 3D

Figure 4. Masked sherds showing detected blobs from each method

From the PCA, the first three principal components explained about 84% of the variation in the data set, with the first 5 exceeding 90% explanation.Because the sample numbers were so low, and by extension the numbers of samples in each group were extremely low, the measures of fit did not work until we scaled back to only use the first three principal components in the function, although not strictly advisable, the results suggest that our visual groupings are justified.

Future Plans

Expand variables: add more observations (e.g. shape, orientation etc.)

Expand to larger datasets: First, here in Seattle with the limited number of sherds we have in storage, but eventually merging with the field school in Spain.

Incorporate a machine learning algorithm: this will be trained using existing ceramic classifications and we anticipate that it will be able to provide a baseline comparison for archaeologists that may lack classification expertise

References and Acknowledgements

[1] Winther-Jacobsen, K. (2010). From Pots to People: A ceramic approach to the archaeological interpretation of ploughsoil assemblages in Late Roman Cyprus. (39)
[2] Stienstra, P. (1986). Systematic Macroscopic Description of the Texture and Composition of Ancient Pottery- Some Basic Methods. *Newsletter of the Department of Pottery Technology*, *4*, 29–48.
[3] Jolliffe Ian T. and Cadima Jorge. (2016). Principal component analysis: a review and recent developments *Phil. Trans. R. Soc. A.*3742015020220150202
[4] Baxter, M. (2003). *Statistics in archaeology* (Arnold applications of statistics series). London : New York: Arnold ; Distributed in the United States of America by Oxford University Press.