Efficient object classification using Euler Characteristic

Erik Amézquita[a], Antonio Rieser[b], Mario Canul Ku[c], Rogelio Hasimoto[c],
Salvador Ruiz Correa[d], Diego Jiménez Badillo[e]
[a] **Math Department, Universidad de Guanajuato**, [b] CONACYT - Centro de Investigación en Matemáticas,
[c] Centro de Investigación en Matemáticas, [d] Instituto Potosino de Investigación Científica y Tecnológica,
[e] Instituto Nacional de Antropología e Historia
erik.amezquita@cimat.mx

Question

Can topology tell us something about the origins of pre-Columbian masks?

General methodology

- The classification algorithm is based on computing the **Euler Characteristic Graph** of each object as suggested in [1].
- We introduced the *extended ECG*, concatenating several ECGs from the same artifact.

Consider a n-dimensional object $X = (V_0, V_1, \ldots, V_n)$ and its Euler Characteristic (EC):

$$\chi = \sum_{k=0}^{n} (-1)^k |V_k|$$

No. of k-dimensional cells

Then fix a vertex filtration function g and extend it to the rest of k-cells:

$$g_k(\{v_0, v_1, \ldots, v_k\}) = \min_{0 \le i \le k} \{g(v_i)\}$$

$g_k : V_k \to [a,b]$ — A k-cell — A fixed function $g : V_0 \to [a,b]$, V_0 set of vertices; $[a,b]$ any interval

The interval $[a,b]$ is divided into T equally-spaced thresholds $a = t_0 < t_1 < t_2 < \ldots < t_T = b$.
Consider **the EC at i-th threshold**:

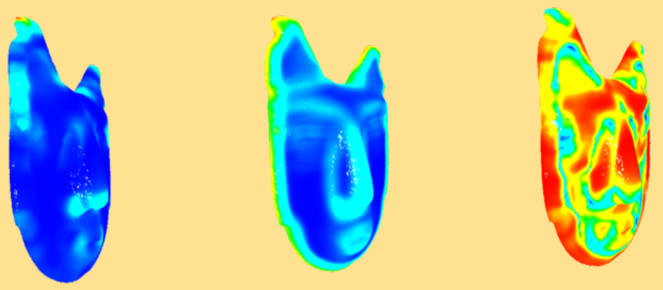
$$\chi_i = \sum_{k=0}^{n} (-1)^k \, |V_k^{(i)}|.$$

No. of k-cells c_k such that $g_k(c_k) \ge t_i$

The **EC Graph (ECG)** is obtained by comparing χ_i vs. t_i.

One particular problem

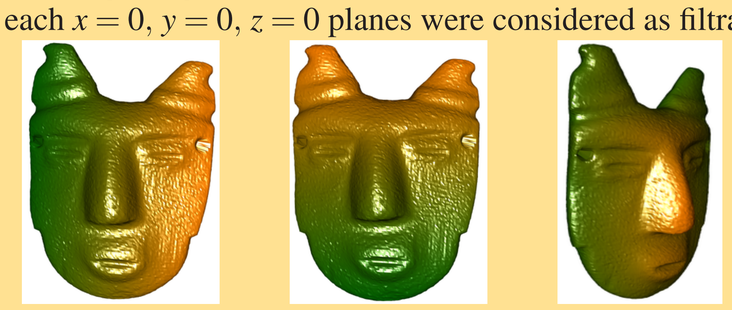
The **goal** was to specifically assort **128 masks** into **9 different groups.** Different filtrations and ECGs were considered.



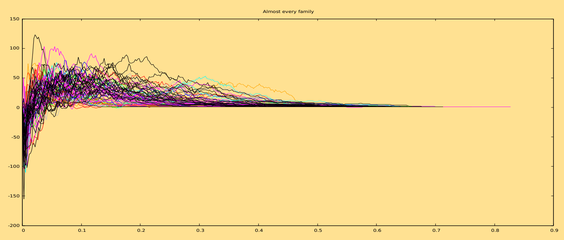
Heatmaps under different filtration functions.
Using the principal curvature values we filtered according to minimum curvature, mean curvature and shape index.

Each mask is embedded in the $[-1,1]^3$ cube, with its mass centered at origin. The projections of each vertex to each $x = 0$, $y = 0$, $z = 0$ planes were considered as filtrations.

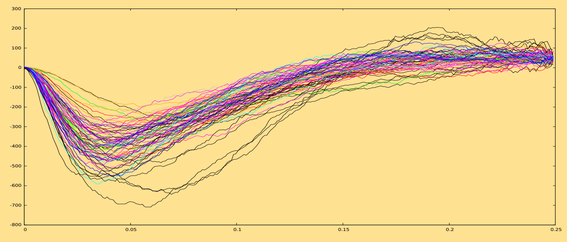
Three different filtrations yield three different ECGs. The three of them are concatenated to get one extended graph for each mask.



All the ECGs for all the masks were computed and plotted on the same plotspace, from which we can infer poor performance based on curvature values.

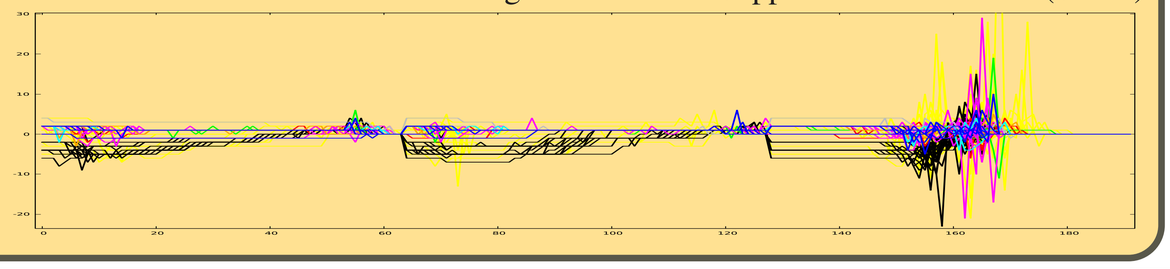


Mean curvature based ECGs for 256 thresholds



Shape index based ECGs for 256 thresholds

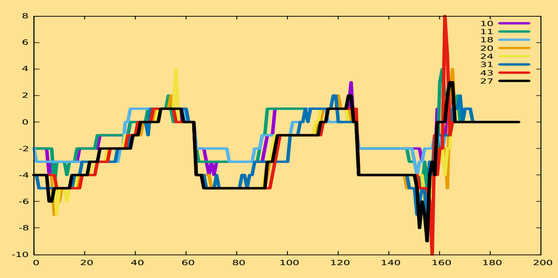
However, our extended ECG obtained by concatenated projections' ECGs gives a more attractive scenario to run a classification algorithm such as Support Vector Machine (SVM.)



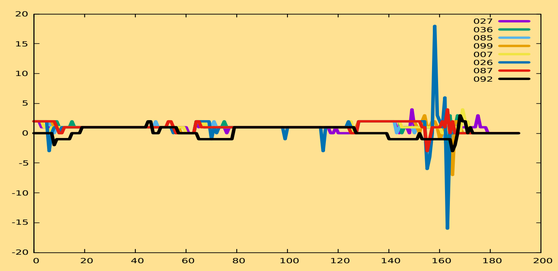
Results

- SVM was run using half of the mask's dataset as training set and the rest of it was used for testing. A **new assortment** was obtained.
- The number of specimens per family was homogenized, since only two out of nine families possess less than 10 items.
- Out of the remaining seven families, 8 items were chosen randomly from each and their projection based ECGs were plotted.
- Different colors refer to different items.

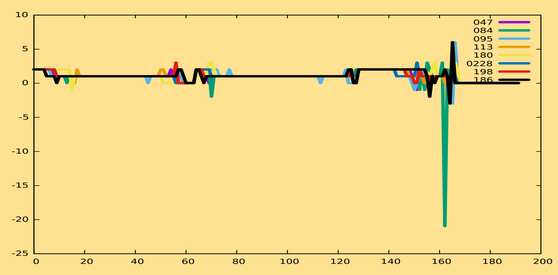
Family 2



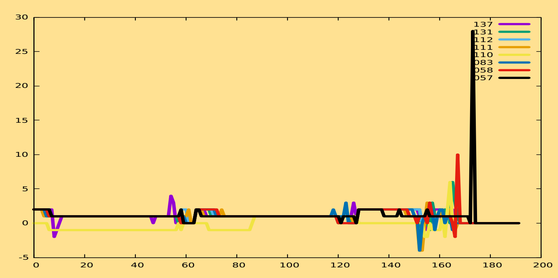
Family 3

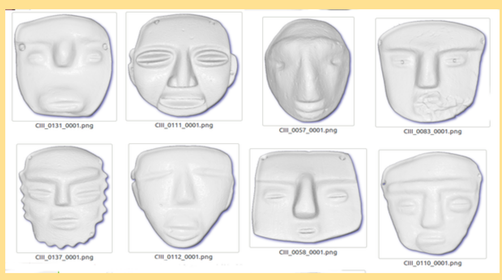


Family 5

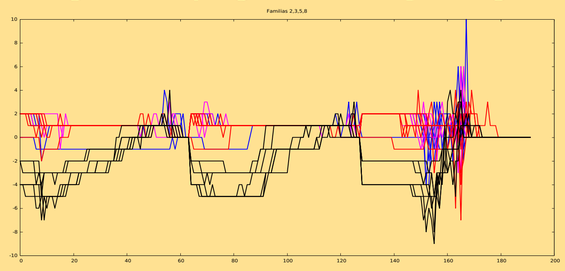


Family 9

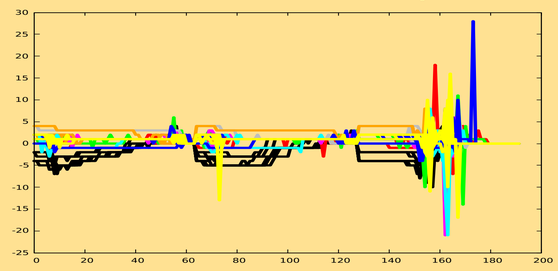




Removing the outlying masks and plotting the same four families simultaneously:



Families 2,3,5,9



New classification

Conclusions

- Answer to the original question: **Yes it can!**
- Archaeologists have manifested some approval towards the first proposed assortment.
- The computation of the ECG associated to a given object, especially if it is based on projections, is a simple algorithm of linear complexity and memory.
- A larger database, with more specimens per family might solve the lack of characterization problem faced throughout the project.
- Outliers suggest that there might not be just nine families in total but more. It would be an interesting problem to determine the appropriate number of families of masks in first place.

Acknowledgments

The authors would like to thank both Instituto Nacional de Antropología e Historia through its project *Desarrollo de Aplicaciones de Computación en Arqueología* and Red Temática CONACYT de Tecnologías Digitales para la Difusión del Patrimonio Cultural for providing the 3D scanned pre-Columbian masks which made possible this project.

References

[1] E. Richardson, M. Weirman Efficient classification using the Euler Characteristic In *Pattern Recognition Letters* Vol.49 pp.99-106, 2014.
[2] C. Chang, C. Lin LIBSVM: A library for support vector machines In *ACM Transactions on Intelligent Systems and Technology* Vol.2, No.3, 2011.