Dominic Delgado
EE 368

Automated Die Study of Greek Coins via Cluster Analysis

Background

One of the tools historians use when trying to understand the economies of ancient civilizations is the die study. Researchers pore over collections of coins from a particular region and attempt to determine which coins were created from the same engraving (die). From the number of dies in use in a region, these researchers can extrapolate about the economic health of the region at that time.

If each die depicted a completely different person or animal, the sorting process would be simple, but often coins will feature the same subject, even if they come from different dies. Since the dies were made by hand, no two dies were identical, but sometimes the differences can be as subtle as the shape of the nose or eye, details on the helmet, or the number of leaves in a palm branch. Since the studies are so time intensive, the corpus of studied coins is much smaller than the number of extant coins. Automating the process could generate much more information about commerce in ancient times.

Problem Statement

Given scanned pages of coin pictures[1], extract the individual coin images and then group them based on distinctive features of the depicted subject.

Algorithm

1. *Image Acquisition* – Binarize the scanned pages, and open with a circular structuring element to eliminate any sparse text. Perform hole filling, if necessary, to complete the coin masks. Then, for each detected coin, extract the pixels from that location and save those pixels as an individual coin image.
2. *Image Registration* – For each coin, detect the foreground of the coin. Register coins so that the foregrounds have the same size and orientation. Crop the edges of the coins so that irregularities in coin edges are not captured in feature extraction.
3. *Feature Extraction* – Apply SIFT, Gabor (or other) wavelets, or some other feature extraction tool to capture edge information on coin surface.
4. *Clustering* – Apply cluster analysis to the features. Initially, in order to determine feasibility, I plan to do this with an algorithm like Lloyd's algorithm (K-means) that requires K to be known, but in order for the tool to be useful in the long run, this would need to use an algorithm that estimates K, such as X-means.

Bibliography

Jain, Anil K. 2010. Data clustering: 50 years beyond K-means. *Pattern Recogn. Lett.* 31, 8 (June 2010), 651-666. DOI=http://dx.doi.org/10.1016/j.patrec.2009.09.011

Liu, Chengjun and H. Wechsler. 2002. Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. *Trans. Img. Proc.* 11, 4 (April 2002), 467-476. DOI=http://dx.doi.org/10.1109/TIP.2002.999679

Pelleg, Dan and Andrew W. Moore. 2000. X-means: Extending K-means with Efficient Estimation of the Number of Clusters. In *Proceedings of the Seventeenth International Conference on Machine Learning* (ICML '00), Pat Langley (Ed.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 727-734.

---

[1] In order to measure the effectiveness of the algorithm, we must use labeled data (i.e. past die studies). Die studies tend to come in book form only.