EUROGRAPHICS Workshop on Graphics and Cultural Heritage (2014)
R. Klein and P. Santos (Editors)

A Large-Scale Evaluation of Correspondence-Based Coin Classification on Roman Republican Coinage

Sebastian Zambanini and Martin Kampel

Computer Vision Lab, Vienna University of Technology, Austria

1. Introduction

While being of great use for the numismatic community, the problem of image-based ancient coin classification has shown to be a challenging problem in the past [KZ08, Ara10, ZK12, ZKK14]. This is caused by the high number of classes, the low number of available samples per class, as well as the general conditions of ancient coins. Another critical problem is the availability of training data. Learning-based methods rely on large amounts of training data to capture the variability within classes, and it is shown in [ZKK14] that the learning-based method proposed in [Ara10] is heavily affected by a low number of training samples. Hence, the method proposed in [ZKK14] is based on a coin-to-coin similarity metric from matched local features and thus suffers less from the training data problem. Accordingly, the method clearly outperforms all previously proposed methods for ancient coin classification [KZ08, Ara10, ZK12] when only one reference coin is available per class.
Although the results in [ZKK14] show the superiority of the proposed method, they are based on a 60-class dataset and it is unclear how the classification performance is influenced by much larger number of classes. In this work, we aim to fill this gap by empirically evaluating the method's performance by means of a dataset consisting of 418 coin classes.

2. Experiments

Dataset: The evaluation dataset consists of 418 classes of Roman Republican coinage where each class is represented by two coin specimens. The images were gathered from the coin collection of the *Museum of Fine Arts, Vienna*, and all classes with at least two available coins have been included. The source collection belongs to the 5 largest collections in the world and hence the 418 classes can be seen as a cross section of the most common classes of the over 1900 (including all subclasses) defined in [Cra74]. For each specimen, single images of the obverse (front) and reverse (back) side are available.

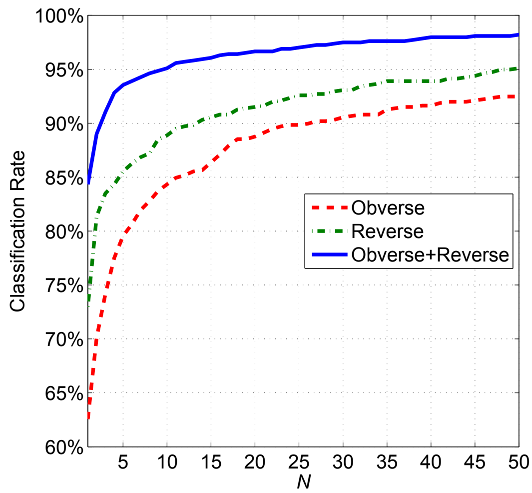

Figure 1: *Classification rate for the top N classes.*

Evaluation Strategy: Each of the 836 coin images in the dataset is classified in an exemplar-based manner by means of the method presented in [ZKK14]. In detail, the dataset is split in two halves and in each half one class is represented by one single coin exemplar. The coins of the first half are compared to all coins of the second half and vice versa. The obtained 418 similarities per input coin image are sorted in order to determine the rank of the correct class. We evaluate the classification performance separately for the obverse and reverse coin sides. Additionally, the improvement of a combined average similarity from both sides is investigated.

Results and Discussion: The classification rates on the evaluation dataset are plotted in Figure 1. The plot shows the respective percentage of coins where the correct class is within the top N similarities. First of all, like in a previous study [ZK12], it can be observed that a classification based on the coins' reverse side has a better performance than a classification based on the obverse sides. This is caused by the higher variation of reverse side motives compared to ob-

DOI: 10.2312/gch.20141326

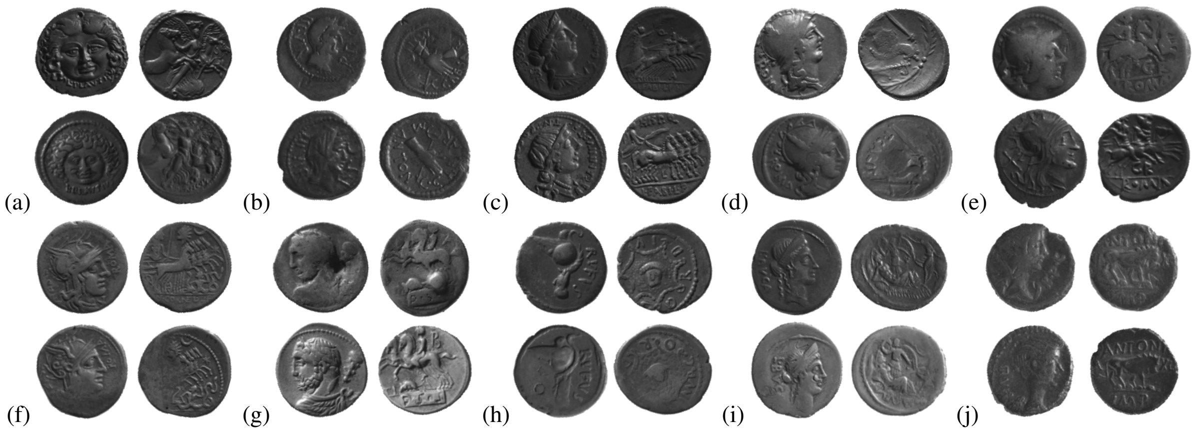

Figure 2: *(a)-(e) Examples for misclassifications where the correct class is not ranked among the top 50 similarities, (f)-(j) examples for correct classifications.*

verse side motives, which typically show heads of mythological or historical persons, and leads to $\sim 10\%$ higher performance (73.0% vs. 62.6% for $N = 1$). Combining the obverse and reverse side similarities boosts the classification rate by another $\sim 10\%$ (84.3% for $N = 1$). It can also be spotted in Figure 1a that there is a comparatively high increase in classification rates from $N = 1$ to $N = 5$: for the combined method the correct class is among the top 5 similarities in 93.5% of the cases. This shows that due to the high number of classes with low inter-class distances an exact classification is challenging, but the ranking is sensitive to the relevant class.

Nevertheless, the curve of Figure 1 is flattened more and more for higher values of N which indicates that there are particular coins where the classification goes completely wrong and the correct class is not ranked properly. For such cases the similarity to the correct class is not considerably higher or even lower than the mean similarity to all incorrect classes. For instance, even with the combined method for 15 coins (i.e. 1.8%) the correct class is not contained in the top 50 similarities. Some of these substantial misclassifications are shown in Figure 2a-e to exemplify their diverse causes. One source of error are scale differences between coin motives which are not sufficiently compensated by the scale normalization induced by the segmentation step. As the local features are extracted with a fixed scale, the correspondence search is disturbed by scale changes which is most evidently seen in the obverse side heads shown in Figure 2a-b. Strong changes in motive appearance (e.g. the chariot of Figure 2c) and abrasions (Figure 2d-e) are further causes for misclassifications. Nevertheless, the correctly classified coin classes shown in Figure 2f-j also exhibit abrasions and motive variations which show that the method is generally able to cope with these types of image variations.

Compared to the original results [ZKK14], which were obtained on 60 classes of reverse side coins, on our dataset a similar classification rate is achieved (73.0% vs. 72.7% in the original evaluation). The dataset of [ZKK14] consists of coins from different sources, whereas all images of the presented dataset come from one single source. This facilitates the classification process, as illumination changes between coin images can be assumed to be much less due to the invariable image acquisition setup. The particular challenge of the dataset is the high number of classes and it is shown that increasing the number of classes has no drastic effect on classification performance.

Notably, in our case the coin reference images have been randomly chosen and obviously the conditions of the reference coins are correlated to the classification performance. Therefore, it can be assumed that in a practical scenario the classification rate can be further improved by a more thorough selection of reference coins.

Acknowledgements: The authors would like to thank Klaus Vondrovec from the Museum of Fine Arts, Vienna, for providing the test images. This research has been supported by the Austrian Science Fund (FWF) under the grant TRP140-N23-2010 (ILAC).

References

[Ara10] ARANDJELOVIĆ O.: Automatic attribution of ancient roman imperial coins. In *Proc. of CVPR* (2010), pp. 1728–1734. 1

[Cra74] CRAWFORD M. H.: *Roman Republican Coinage*. Cambridge University Press, 1974. 1

[KZ08] KAMPEL M., ZAHARIEVA M.: Recognizing ancient coins based on local features. *Proc. of ISVC* (2008), 11–22. 1

[ZK12] ZAMBANINI S., KAMPEL M.: Coarse-to-fine correspondence search for classifying ancient coins. In *ACCV 2012 eHeritage Workshop* (2012), pp. 25–36. 1

[ZKK14] ZAMBANINI S., KAVELAR A., KAMPEL M.: Classifying ancient coins by local feature matching and pairwise geometric consistency evaluation. In *Proc. of ICPR* (2014). (in print). 1, 2