Improving Ancient Roman Coin Recognition with Alignment and Spatial Encoding

Jongpil Kim(✉) and Vladimir Pavlovic

Department of Computer Science, Rutgers University, Piscataway, NJ 08854, USA
jpkim@cs.rutgers.edu

Abstract. Roman coins play an important role to understand the Roman empire because they convey rich information about key historical events of the time. Moreover, as large amounts of coins are daily traded over the Internet, it becomes necessary to develop automatic coin recognition systems to prevent illegal trades. In this paper, we describe a new large annotated database of over 2800 Roman coin images and propose an effective automated system for recognition of coins that leverages this new coin image set. As the use of succinct spatial-appearance relationships is critical for accurate coin recognition, we suggest two competing methods, adapted for the coin domain, to accomplish this task.

Keywords: Recognition · Detection · Coin recognition

1 Introduction

A coin is usually a flat piece of metal issued by governmental authority as a medium of exchange. It has been produced in large quantities to facilitate trade from the ancient history to the present. Along with the trading purpose, the Roman empire knew how to effectively use the coin as their political propaganda. The ancient Roman coins were widely used to convey the achievements of Roman emperors to the public. They also served to spread messages of changing policies or merits through the empire. By engraving portraits on the coins, the Roman emperors also could show themselves to the entire empire. In short, the coins were the newspaper of the Roman empire. In this way, Roman coins are always connected to historical events and Roman imperial propaganda. Therefore, understanding the ancient Roman coins could serve as references to understand the Roman empire.

Because the coin market is very active, many coins are traded every day, mostly over the Internet [1]. But ancient coins are also becoming subject to a very large illicit trade [21]. The traditional way to detect illegal traffic of ancient coins is to manually search catalogues, dealers or internet by the authorities. But the manual process has limitations and is too slow to cover all trade. Therefore, there is a need to develop both reliable and automatic methods to recognize the coins.

L. Agapito et al. (Eds.): ECCV 2014 Workshops, Part I, LNCS 8925, pp. 149–164, 2015.
DOI: 10.1007/978-3-319-16178-5_10

(a) Vespasian (b) Vitellius

Fig. 1. An example of Inter-class similarity in the ancient Roman coins. Vespasian looks similar to Vitellius.

There are tens of thousands of typologies that could be used to classify Roman coins [1][3]. Therefore those who do not have knowledge and experience cannot classify them without the help of experts or automatic classifiers. In this paper, we focus on the recognition of the Roman emperors on the Roman imperial coins. Specifically, for a given coin image, we propose an automatic method to recognize who is on the coin.

Inter-class similarity and intra-class similarity are two challenges to recognize the ancient Roman coins. For the inter-class similarity, different emperors share similar appearance as shown in Figure 1 and Figure 10. There are several reasons for the similar appearance: familiar relationship, engraver's lack of knowledge for the emperor's image or abstraction, or using the same template for different emperors. Another aspect of the coin recognition challenge is the intra-class dissimilarity as shown in Figure 2. There may exist a large variation within the same class. On a very basic level, the direction of the emperor's face varies over the coins: some emperors look left and the others look right without any specific rule as shown in Figure 2.

Several works [15], [16], [1], [9], [21], [22], [17] have proposed to recognize the coins using computer vision techniques. In general, they represent the coin image as low level visual features such as SIFT [14] and perform the recognition using the k-nearest neighbor method or the support vector machine. Among them, Arandjelović [1] introduces a new type of feature called directional kernel which captures geographical information between interest points. On the other hand, Zambanini *et al.* [17] employs the spatial pyramid models [13] to capture the structure of the coin. However, the previous approaches do not explicitly facilitate the use of the spatial structure of the coin such as the location of the face on the coin.

In this paper, we address the problem of automatically recognizing ancient Roman coins, while leveraging their spatial structure and without specifically focusing on the understanding of textual transcripts on coins. The ancient Roman coins have regular structure: the coin is round, the location of the emperor is roughly at the center of the coin and the emperors share common aspects across different coins. However, coins of the same emperor also exhibit large variations. Some of these variations are due to the differences in the coin material and

Fig. 2. The emperors on the five coins are the same, Nero. But there is variations on the shapes. In particular, one face looks left while the others look right.

diverse state of coin degradation. Others are due to the differences in which the same emperor was depicted by different coin creators. These intra-class variability aspects make the task of recognizing the coins very challenging. An additional challenge, as we demonstrate in this work, comes from the fact that the visual coin appearance gives rise to vastly different feature statistics compared to those of traditional face recognition tasks. To surmount these challenges we propose a framework to simultaneously leverage the consistencies in the coin structure and local appearance to improve the recognition accuracy.

To this end, we investigate two approaches: a method based on discriminative deformable part models (DPM) specifically adjusted to the coin domain through the use of polar coordinate representations and the Fisher vector with spatial-appearance encoding. The model using DPM first detects the face of the emperor on the coin and uses the detected location to build a spatial pyramid. The Fisher vector based model directly encodes the spatial information in its representation. The use of both representations allows the recovery of consistent patterns that characterize different Roman emperors despite the outlined intra-class differences. We use the support vector machine (SVM) to train the models. We also introduce a new large annotated database of Roman coins, consisting of over 2800 pieces made of different materials, depicting appearances of 15 Roman emperors. This new dataset allows us to establish the performance advantages of the proposed approaches compared with more traditional methods.

Contribution of this paper are twofold: 1) we have collected a new ancient Roman Imperial coin dataset where all the coins are annotated and consist of high-quality images; and 2) to leverage the new image set we introduce new baseline techniques for the coin classification using the deformable part model and the Fisher vector with spatial-appearance encoding. We believe that the baseline techniques will provide a benchmark for the future coin recognition problem.

This paper is organized as follows: Related work is summarized in Section 2. In Section 3, we explain our proposed method to recognize the ancient Roman coin. Then, we explain the coin dataset and show experimental results in Section 4. Lastly, we make conclusions in Section 5.

2 Related Work

Several methods have been proposed to recognize and analyze coins using computer vision techniques. Among those methods, gradient information based

approaches [15], [16] and eigenspace decomposition based approaches [9] were proposed to recognize modern coins. But none of them are adequate for the ancient coin classification because the ancient coins are too often in very poor conditions, common recognition algorithms can easily fail [21]. Therefore, we do not consider those approaches in this paper. In [10], SIFT descriptors [14] are used to obtain 90% classification accuracy for 390 coin images where there are only 3 classes. A directional kernel to consider orientations of pixels was proposed in [1] for ancient Roman coin classification. However, we found that directional histograms used in [1] performed worse than the proposed method on our dataest. We will examine the results of the directional kernel method in Section 4.2.

Recently, a bag-of-words approach to coin recognition with standard spatial pyramid models was proposed in [17]. In [17], rectangular spatial tiling, log-polar spatial tiling and circular spatial tiling methods were used to recognize the ancient coins. However, they did not align the coin images. In this paper, we propose a new method to align the coin images by detecting the face on the coin. By comparing to the standard spatial pyramid model approach, we will show that the alignment is crucial to improve the recognition accuracy.

The coin recognition problem can be considered as the face recognition in terms of recognizing an Emperor's face on the coin. Many methods have been developed for the recognition of real face images. However, the use of such methods faces significant challenges when applied to ancient coins. In terms of its gradient/edge content, a typical critical feature used in face recognition, most ancient coins display vastly different statistics from photographed faces. This aspect is further aggregated by the fact that many coins are old, worn-out and damaged. For example, Figure 3 shows different HoG [4] distributions between the coins and the real faces. Please note that coins show random distributions while all the human faces share the common pattern.

Tzimiropoulos *et al.* [19] proposed a method to learn a subspace from image gradient orientations (IGO subspace learning) for appearance-based face recognition that was shown to be very robust to different types of image noise. The advantage of using the IGO subspace learning algorithm is that the cosine distance measure of the algorithm can cancel out outliers or noise caused by occlusion or illumination changes. However, the IGO-algorithm is very sensitive to the alignment, requiring exactly aligned images. We will examine the performance of the IGO-algorithm on the ancient Roman coin recognition problem in Section 4.

Coin recognition methods to read legends of Roman coin were proposed in [2] [11]. The legend provides rich contextual information about the coin such as issuer, mint date and emperor. Therefore, reading the legend means that we know almost everything about the coin including the identification of the emperor. In this sense, the coin recognition is different from the face recognition where there is no contextual information. However, because the letters in the legend are prone to be damaged as shown in Figure 1 and Figure 10, we can extract information from the legend only if we are given a very well-preserved coin. Therefore, we focus on the face of the emperor and implicitly use the legend as we extract features from the legend area without explicitly recognizing it.

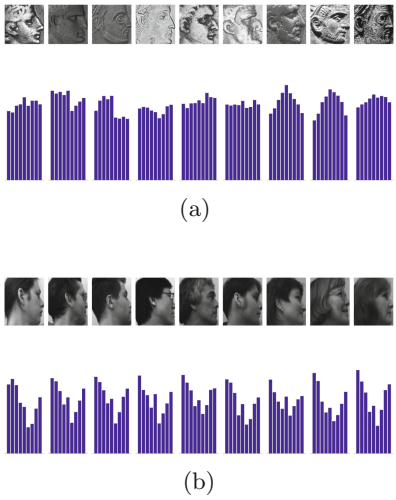


Fig. 3. HoG Distributions for coin and real face images. For each image, we extract the HoG descriptors with 9 bins and average them to build a histogram. TOP: 9 coin images. BOTTOM: 9 profile images. The real faces show regular patterns while the coins show different distributions.

3 A Spatial-Appearance Coin Recognition System

In this section, we explain ancient Roman coin recognition methods. First, we explain the standard bag-of-words approach with the spatial pyramid model. Second, we explain a new framework to recognize the ancient coin by employing the deformable part model [7] to locate the face of the emperor. Third, we introduce an alternative method for coin classification which directly encodes the spatial information based on the Fisher vector method.

3.1 Spatial Pyramid Model Approach

We use the bag-of-words (BoW) model to represent the coin images as visual histograms. In the BoW model, we select a set of key points on an input image, extract descriptors from the set of the points, quantize the descriptors to a visual codebook and represent the input image as a histogram of the codewords in the codebook.

Several strategies including dense regular grid [6] and difference-of-Gaussian (DoG) peaks [14] can be used to find the key points on the image. Among them, the dense grid sampling method shows better performance than the others so we choose it to find the key points in our system. On the grid points of the image, we extract SIFT descriptors [14] . k-means clustering method is then used to build the visual codebook and the image is represented as a histogram of the visual words from the codebook.

Fig. 4. An example of the spatial pyramid based on the polar coordinate system. An image is divided into 16 non-overlapping regions.

One problem of this approach is that all the key points are equally handled and the spatial locations of the points are ignored. Therefore, the dense sampling method makes no use of information of the coin structure. To overcome this limitation, we follow the spatial pyramid models widely used in computer vision area [13], [5]. As in [17], we use the polar coordinate system to take an advantage of round shapes of the coin and the face of the emperor. In the polar coordinate system, we place the center of the system at the mass center of the coin and an image is divided into 16 non-overlapping wedge shape regions as shown in Figure 4. The area of the outer circle is equal to the area of the coin and the radius of the inner circle is 60% of the radius of the outer circle.

But the polar system is also sensitive to the origin placement. In the next section, we employ the deformable part model to address the finding of the origin placement

3.2 Strongly-supervised Recognition using Deformable Part Model

In the ancient Roman coins, the shape of the face is a characteristic differentiating a Roman emperor from the other emperors. As coins are old and worn-out the location of the face varies across the coins as shown in Figure 5. In this paper, we employ the deformable part model (DPM) [7] to align the coin image by locating the face of the emperor.

To train DPM with the emperor's face, we define a bounding box of the face as depicted in Figure 5. Four outermost points to draw the bound box are as follows:

- "forehead": a point where the face and the hair meet
- "chin": a point where the chin and the neck meet
- "ear": the left most point of the ear
- "nose": the tip of the nose.

They are chosen so that the face areas inside the box agree all over the emperors. Negative images are collected from reverse images because they contain symbols without the face of the Roman emperor. After training, DPM will automatically detect the face location for a new coin image.

To align the coin image we place the center of the polar coordinate system onto the center of the detected face area by DPM. In this way, each region defined by the polar coordinate system contains consistent area of the coin. We

then extract visual features on sub-regions of the polar coordinate system, build a histogram on each region and concatenate them to represent the image.

3.3 Coin Classification Using Fisher Vector

The coin recognition problem can be considered as the fine-grained classification problem as all the images belong to the same super class of 'coin' in this paper. In the fine-grained classification, the distribution of the gradients are often similar as the shapes of the coins look similar to each other. Therefore, Fisher vectors are able to describe the subtle changes in the gradients since they are designed to capture both the first order and the second order statistics of the gradient distribution [8].

Let each image patch be presented as $(\mathbf{x}, \mathbf{l})$ where $\mathbf{x}$ is a vector of visual descriptors of the patch and $\mathbf{l}$ the 2-D location of the patch. We assume that the visual descriptor and the location are independent from each other. We also assume that the patch is generated from a Gaussian mixture model (GMM). As in [12], we formulate the model as follows:

$$\begin{aligned} p(\mathbf{x}, \mathbf{l}) &= \sum_k \pi_k \cdot p(\mathbf{x}, \mathbf{l}; \Sigma_k^V, \Sigma_k^L, \mu_k^V, \mu_k^L) \\ &= \sum_k \pi_k \cdot p(\mathbf{x}; \Sigma_k^V, \mu_k^V) \cdot p(\mathbf{l}; \Sigma_k^L, \mu_k^L), \end{aligned} \tag{1}$$

π_k is a prior probability for the kth component, Σ_k^V, μ_k^V are means and covariances for the visual descriptors, Σ_k^L, μ_k^L mean and covariance for the location, and

$$p(\mathbf{x}; \Sigma_k^V, \mu_k^V) \sim \mathcal{N}(\mathbf{x}; \Sigma_k^V, \mu_k^V) \tag{2}$$

$$p(\mathbf{l}; \Sigma_k^L, \mu_k^L) \sim \mathcal{N}(\mathbf{l}; \Sigma_k^L, \mu_k^L). \tag{3}$$

Then, the gradient of $\ln p(\mathbf{x}, \mathbf{l})$ with respect to all parameters becomes the Fisher vector to represent the patch $(\mathbf{x}, \mathbf{l})$. An image representation is obtained by averaging the gradients over all patches in the image. Let $I = \{\mathbf{f}_1, \ldots, \mathbf{f}_N\}$ be a set of D dimensional feature vectors extracted from an image where $\mathbf{f}^{D\times 1} = [\mathbf{x}^\top, \mathbf{l}^\top]^\top$. Then we have

$$u_{jk} = \frac{1}{N\sqrt{\pi_k}} \sum_{i=1}^{N} q_{ik} \frac{f_{ji} - \mu_{jk}}{\sigma_j}, \tag{4}$$

$$v_{jk} = \frac{1}{N\sqrt{2\pi_k}} \sum_{i=1}^{N} q_{ik} \left[\left(\frac{f_{jk} - \mu_{jk}}{\sigma_{jk}} \right)^2 - 1 \right], \quad (j = 1, \ldots, D) \tag{5}$$

where σ_{jk} is the jth diagonal entry of Σ_k and

$$q_{ik} = \frac{p(\mathbf{f}_i; \Theta_k)}{\sum_t p(\mathbf{f}_i; \Theta_t)} \qquad \left(\Theta_k = \Sigma_k^V, \Sigma_k^L, \mu_k^V, \mu_k^L\right).$$

The Fisher vector of the image I is the stacking of the vectors $\mathbf{u}_k$ and $\mathbf{v}_k$. Typically, the parameters Θ_k are trained by the Gaussian mixture model. In this paper, we use the implementation of [20] to represent the coin image as the Fisher vector.

4 Datasets and Experimentation Results

4.1 Coin Data Collection and Experimental Settings

We collect ancient Roman coin images from a numismatic web site. Each coin has a high resolution image (approximately 350×350 pixels jpeg image). Among the collected coin images, we found that some of the coins are hard to recognize because they are rusty and severely damaged. As we are dealing with the problem of recognizing the Roman emperors, a coin that is severely damaged or hard to recognize who is on the coin is discarded. After removing such worn-out coins, we select emperors who appeared more than 10 times in the dataset. Finally, we arrive at 2815 coins with 15 emperors. The sample images for the 15 emperors are depicted in Figure 10. All images in the dataset are ancient Roman Imperial coins dated from 27 BC to 355 AD. In this paper, we consider only the observe(front) of the Roman coin because the emperor is engraved in the observe and the reverse usually shows various non-face symbols.

The deformable part model (DPM) needs the bounding box information in the training procedure. Therefore, we manually annotate bounding boxes for randomly selected 500 coins. Because the coin has the regular structure, we found that 20% of the whole images is enough to train DPM and produce good detection results. As shown in Figure 5, four outermost points to define the bound box are "forehead", "chin", "ear" and "nose", all of which are easily determined by people.

The SIFT descriptor [14] is used to extract the visual descriptors on the grid points. The grid step size (spacing) is 4. Then we use k-means clustering to build the visual codebook, generating the histograms for the coin images using the visual codebook. In this paper, we set k to 200.

The multi-class SVM is used for training and prediction with the RBF-χ^2 kernel except for the Fisher vector based method. We use the linear SVM for the Fisher vector method because it is known that the Fisher vector performs well with simple linear classifiers [18]. All parameters are determined by the 5-fold cross validation. For the evaluation, we randomly partition the coin dataset into 5 equal size subgroups. Each subgroup keeps the same emperor ratio as the total coin dataset. Then, we use 4 subgroups as training data and 1 subgroup as test data. Experiments are repeated 5 times so that each of 5 subgroups becomes the test data, and we report the average of the 5 outputs as a final output. We calculate the average number of correctly classified test samples as the evaluation measure.

Fig. 5. Examples of DPM detection results. The red box represents the ground truth location of the face and and the blue box shows the detected face location by DPM. In most cases, DPM is able to detect the face area. The average accuracy of DPM is 87.18%.

4.2 Experimental Results

We first examine the performance of the deformable part model[1] on the coin dataset. Figure 5 depicts the results of DPM for randomly selected 4 coin images. The red box represents the manually annotated face area of the emperor and the detection area is given by the blue box. We can observe that DPM finds the face areas accurately. To quantitatively evaluate the performance, we use the number of overlapped pixels divided by the size of the ground truth face area as the evaluation measure. We average measured values over the testset and have 0.8718. The most difficult part to detect is the forehead point where the face and the hair meet as shown in the first image of Figure 5. We will use the detected face area in the next part to recognize the coin.

Next, we measure the performance of various method for the coin recognition. To set the baseline performance, we use the standard bag-of-words method without using the spatial pyramid model (`NoSpatial`). In the second method, similar to [17] we use the bag-of-words method with the polar coordinate system centered at the mass center of the coin (`Polar`). `Polar` presumes that the all the coin images are aligned along the mass center of the coin. Neither `NoSpatial` nor `Polar` takes the alignment of the coin into account while recognizing the coin. We also examined the performance of the standard spatial pyramid with the rectangular tiles and found that it performed worse than `Polar`. We believe that the polar coordinate system is more appropriate to capture the round shape of the coin than the rectangular tiles.

However, the location of the Roman emperor's face varies across the coin dataset. Therefore, we use the center of the face detected by DPM to align the coin images as many methods are sensitive to the alignment. To take the advantage of DPM, we construct the spatial pyramid model with the polar coordinate system centered at the center of the detected face by DPM (`DPM-Polar`).

The next method is based on the Fisher vector which encodes the spatial information as described in Section 3.3 (`FV`). Unlike `DPM-Polar`, `FV` does not need additional training procedure but facilitates the spatially local information in the framework of the Fisher vector approach. In this paper, we assume that

[1] We use the implementation available at http://people.cs.uchicago.edu/~rbg/latent-release5/.

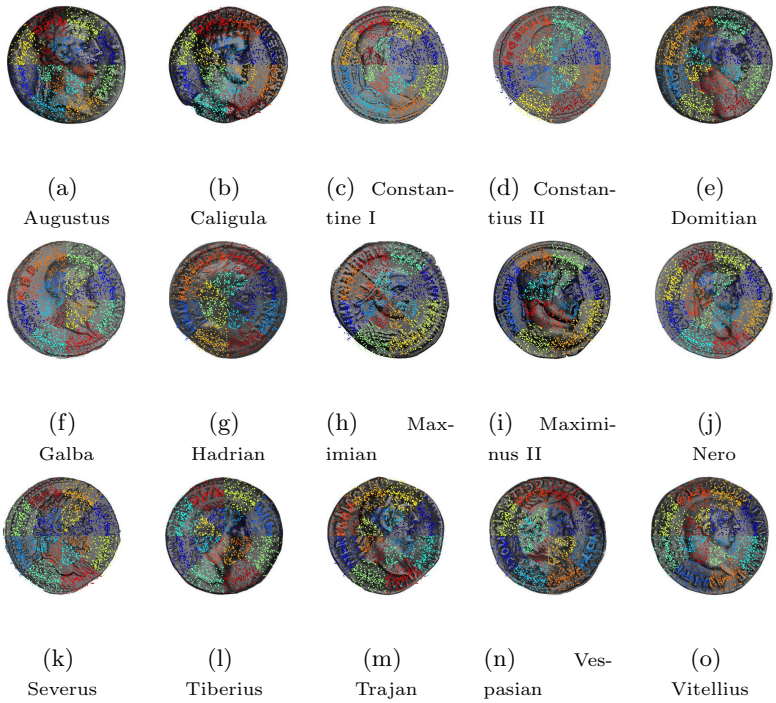


Fig. 6. Discriminative Regions for 15 emperors. Red for more discriminative and blue for less discriminative.

covariance matrix is the identity matrix and the mean vectors are generated by k-means clustering method because not only it is computationally efficient but also it shows better results than using means and covariance matrices from the Gaussian mixture model. We think that the Gaussian mixture model is rather over-fitted on the coin dataset. We do not use the Fisher vector representation with the spatial pyramid because it has not improved the recognition accuracy. In this paper, we set the number of mixtures in the Fisher vectors to 2000.

Then, we use the IGO subspace learning method [19] to examine how well the state-of-the art human face recognition method works for the coin recognition problem (`IGO-PCA`). We choose `IGO-PCA` because it was shown to be robust to different types of image noise, which often occur in the coin images. We align and crop the faces along the results of DPM to train `IGO-PCA`.

Regarding the previous coin classification method, we compare to the directional kernel approach [1](`DK`) which facilitates the spatial structure of the coin by considering the relationship of the gradient directions between two points on the coin.

The recognition accuracies are summarized in Table 1. `NoSpatial` shows worse performance than the others considering the spatial structure as it does not take the spatial information into account. On the other hand, `Polar` significantly

Table 1. Recognition accuracies for various methods

	NoSpatial	Polar	DPM-Polar	FV	IGO-PCA	DK [1]
Acc.	65.77%±1.24	83.48%±1.12	**85.93**%± 0.64	85.36%±0.72	67.42%±0.0097	33.0(±3.0)

improves the recognition accuracy over NoSpatial. The improvement implies that Polar is able to the advantage of the regular structure of the coin.

DPM-Polar shows the best performance among the competing methods because it facilitate the additional spatial information provided by DPM. FV performs comparable to DPM-Polar as it directly encodes the location information without defining spatial pyramids. This result implies that FV provides rich information about the structure of the coin.

IGO-PCA performs slightly better than NoSpatial and at least 16% worse than the others. The poor performance of IGO-PCA can be explained that that coin recognition is totally different from human face recognition and it is difficult to simply apply methods of human face recognition to the ancient Roman coin recognition problem.

DK shows the worst performance on our coin dataset. DK depends on the difference-of-Gaussian (DoG) sampling method to find a set of interest points on the coin image. To combine the visual feature (*e.g.*, SIFT) and the spatial information, it takes the orientation of the interest points into account to build the directional histogram. Therefore, DK becomes sensitive to the orientation of the pixel on the coin image. However, because not only coins in our dataset are old and sometimes damaged but also coin images are taken in different illumination conditions (this is typical in the coin images), it is likely that the orientations of the interest points mislead the classification, explaining the poor performance of DK. On the other hand, as the proposed method uses the uniform sampling on the grid and facilitates the location of the pixels instead of their orientations , it can reduce the effect of the noise of the coin image and therefore it is more robust than DK on the coin dataset.

Discriminative Region Analysis. To determine which area is more discriminative than the others in terms of the recognition, we use the Fisher vector representation described above. Because each component of the Gaussian mixture model (GMM) has the location information in the proposed method, we can project the component of GMM on the 2-D plane according to its location. Then, the importance of each component is determined by the weights derived from the linear SVM model. For the visualization of the discriminative region, we divide the coin area into 16 regions so that the sum of SVM weights of all components in a region represents the importance of the region.

Figure 6 depicts the experimental results. Red color represents more discriminative, blue less discriminative. In general, the upper head area is more discriminative than the other regions because it covers the hair style which is one of the important characteristics of the Roman Emperor as shown in Figure 10. On the other hand, upper right and lower left are less discriminative because they do not

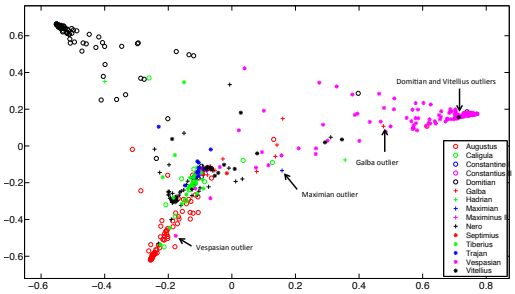


Fig. 7. Projection of the coin images on the 2-D plane using PCA. Each symbol corresponds to one image. A black arrow points an outlier which is far from the same class and/or close to another class.

contain the Emperor's face and the letters on those areas are not consistent over the coin dataset which we have collected. For the Emperors who have specific beard styles such as Maximian (h), Maximinus II (i) and Severus (k), `chin` and `neck` areas are discriminative. The analysis implies that the hair and beard style can provide guidance to discriminate the ancient Roman coins to those who do not have much knowledge on the ancient Roman coins.

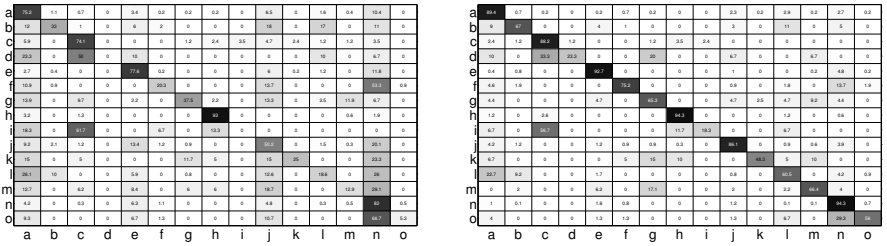
Outlier Detection. To see outliers in the coin classification, we collect confidence values of the coin images by running the trained multi-class SVM. Therefore, each coin image is represented as a vector of the confidence values from SVM. Then, we study the manifold of this space of confidences. We do this by running the principal component analysis (PCA) on the confidence values and choose the top 2 principal components. This process allows us to map the coin images on the 2-D plane along the top 2 principal components so we can evaluate the relationship between two emperors with respect to the distance between them.

Figure 7 depicts the projection of the coin images on the 2-D plane. We can find that several emperors form groups together. Outliers can be determined based on the distance to the center of the group. In addition to identifying outliers, we can also judge which emperors are more similar to each other, which ones may be more difficult to distinguish. For evaluation, we select 5 images denoted by the black arrows in Figure 8.

Figure 8a (resp., Figure 8b) depicts Domitian (resp., Vitellius) which is far from the group of Domitian (resp., Vitellius) and close to Vespasian in the projection plane. Both images look like Vespasian in terms of the appearance. In Figure 8d, Galba which is closer to the group of Vespasian than the group of Galba looks like Vespasian as it loses details because of the damage on the coin. One can find that Maximian in Figure 8e looks very different from the ordinary appearance as shown in Figure 10h. Finally, because of the severe damage, Vespasian in Figure 8f is far from the others, making it the outlier.

(a) Domitian (b) Vitellius (c) Vespasian (d) Galba (e) Maximian (f) Vespasian

Fig. 8. (a),(b): Outliers for Domitian and Vitellius which look similar to Vespasian (c). (d) Outlier for Galba which looks similar to Vespasian because of the damage on the coin. (e) Outlier for Maximian which looks totally different from the ordinary shape as shown in Figure (10h). (f) Outlier for Vespasian which is damaged and far from the other Vespasian images.



(a) Confusion matrix for `NoSpatial` (b) Confusion matrix for `DPM`

Fig. 9. Confusion matrices for `NoSpatial` and `DPM`. `DPM` has improved the recognition accuracies across all the emperors. (Indices in the table correspond to indices in Figure 10).

Confusion Matrices. The confusion matrices for `NoSpatial` and `DPM` are depicted in Figure 9. One can find that `DPM` has improved the recognition accuracies across all the emperors. Constantius II (d) and Maximinus II (i) are not recognized using `NoSpatial` because the number of training samples for them is much smaller than the others. However, because of facilitating the spatial information to capture the characteristics of the emperors, `DPM` is able to improve the recognition accuracies for them.

Maximian (h) has the highest recognition accuracy because he has a distinct appearance from the others as shown in Figure 10. On the other hand, the most confused cases for `DPM` occurs between Constantine I(c) and Maximinus II (i) because they share very similar appearance as shown in Figure 10.

(a) Augustus (b) Caligula (c) Constantine I (d) Constantius II (e) Domitian

(f) Galba (g) Hadrian (h) Maximian (i) Maximinus II (j) Nero

(k) Severus (l) Tiberius (m) Trajan (n) Vespasian

(o) Vitellius

Fig. 10. An example observe image of a coin for each of the 15 classes in the dataset

5 Conclusion

We proposed two automatic methods to recognize the ancient Roman coins. The first method employs the deformable part model to align the coin images to improve the recognition accuracy. The second method facilitates the spatial information of the coin by directly encoding the location information. As the first method takes the information of the face location into account, it performs slightly better than the second method. The experiments show that both methods outperform the other methods such as the standard spatial pyramid model and human face recognition method.

The contributions of the paper are collecting a new ancient Roman coin dataset and proposing an automatic framework to recognize the coins where

we employ the state-of-the-art face recognition system and facilitate the spatial information of the coin to improve the recognition accuracy. The coin images are high-resolution (350-by-350 pixels) and the face locations are annotated. While the proposed coin recognition framework is based on the standard methods such as bag-of-words with spatial pyramids, Fisher vectors and DPM, we believe that their use in the context of the ancient coin recognition represents an interesting contribution.

References

1. Arandjelović, O.: Automatic attribution of ancient Roman imperial coins. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2010)
2. Arandjelović, O.: Reading ancient coins: automatically identifying denarii using obverse legend seeded retrieval. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part IV. LNCS, vol. 7575, pp. 317–330. Springer, Heidelberg (2012)
3. Cohen, H.: Description historique des monnaies frappées sous l'empire romain, vol. I-VIII, 2 edn. Paris, France (1880–1892)
4. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2005)
5. Dimitrovski, I., Kocev, D., Loskovska, S., Deroski, S.: Hierarchical annotation of medical images. Pattern Recognition **44**, 2436–2449 (2011)
6. Fei-Fei, L., Perona, P.: A bayesian hierarchical model for learning natural scene categories. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2005)
7. Felzenszwalb, P.F., Girshick, R.B., McAllester, D., Ramanan, D.: Object detection with discriminatively trained part based models. IEEE Transactions on Pattern Analysis and Machine Intelligence **32**, 1627–1645 (2010)
8. Gavves, E., Fernando, B., Snoek, C.G.M., Smeulders, A.W.M., Tuytelaars, T.: Fine-grained categorization by alignments. In: International Conference on Computer Vision (ICCV) (2013)
9. Huber, R., Ramoser, H., Mayer, K., Penz, H., Rubik, M.: Classification of coins using an eigenspace approach. Pattern Recognition Letters (PATREC) **26**, 61–75 (2005)
10. Kampel, M., Zaharieva, M.: Recognizing ancient coins based on local features. In: Bebis, G., Boyle, R., Parvin, B., Koracin, D., Remagnino, P., Porikli, F., Peters, J., Klosowski, J., Arns, L., Chun, Y.K., Rhyne, T.-M., Monroe, L. (eds.) ISVC 2008, Part I. LNCS, vol. 5358, pp. 11–22. Springer, Heidelberg (2008)
11. Kavelar, A., Zambanini, S., Kampel, M.: Word detection applied to images of ancient Roman coins. In: International Conference on Virtual Systems and Multimedia (VSMM), pp. 577–580, September 2012
12. Krapac, J., Verbeek, J., Jurie, F.: Modeling Spatial Layout with Fisher Vectors for Image Categorization. International Journal of Computer Vision (IJCV) (2011)
13. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In: IEEE Computer Vision and Pattern Recognition (CVPR) (2006)
14. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision (IJCV) **60**, 91–110 (2004)

15. van der Maaten, L.J.P., Boon, P.J.: Coin-o-matic: a fast system for reliable coin classification. In: MUSCLE CIS Coin Competition Workshop (2006)
16. Nölle, M., Penz, H., Rubik, M., Mayer, K., Hollaender, I., Granec, R.: Dagobert - a new coin recognition and sorting system. In: Conference on Digital Image Computing: Techniques and Applications (DICTA) (2003)
17. Piater, J., Rodríguez-Sánchez, A.: A Bag of Visual Words Approach for Symbols-Based Coarse-Grained Ancient Coin Classification. CoRR abs/1304.6, 1–8 (2013)
18. Sanchez, J., Perronnin, F., Mensink, T., Verbeek, J.: Image Classification with the Fisher Vector: Theory and Practice. International Journal of Computer Vision (IJCV) **105**, 222–245 (2013)
19. Tzimiropoulos, G., Zafeiriou, S., Pantic, M.: Subspace learning from image gradient orientations. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) **34**, 2454–2466 (2012)
20. Vedaldi, A., Fulkerson, B.: VLFeat: An open and portable library of computer vision algorithms (2008). http://www.vlfeat.org/
21. Zaharieva, M., Kampel, M., Zambanini, S.: Image based recognition of ancient coins. In: Kropatsch, W.G., Kampel, M., Hanbury, A. (eds.) CAIP 2007. LNCS, vol. 4673, pp. 547–554. Springer, Heidelberg (2007)
22. Zambanini, S., Kampel, M.: Coarse-to-fine correspondence search for classifying ancient coins. In: Park, J.-I., Kim, J. (eds.) ACCV Workshops 2012, Part II. LNCS, vol. 7729, pp. 25–36. Springer, Heidelberg (2013)