A ROTATION-INVARIANT BAG OF VISUAL WORDS MODEL FOR SYMBOLS BASED ANCIENT COIN CLASSIFICATION

Hafeez Anwar, Sebastian Zambanini, Martin Kampel

Computer Vision Lab, Institute of Computer Aided Automation
Vienna University of Technology, Vienna, Austria

ABSTRACT

We propose to perform image-based ancient coin classification by recognizing symbols minted on the reverse side of coins. Dense sampling based bag-of-visual-words model is used for symbol recognition. The lack of spatial information in the bag-of-visual-words model degrades symbol recognition rate as the symbols have specific geometric structures. Furthermore, coins can be imaged under various rotations resulting in severely rotated symbols. Therefore we propose a novel bag-of-visual-words model for symbol-based coin classification which accounts for the spatial arrangement of the visual words in a rotation invariant manner. We perform our experiments on images collected from three different sources thus making our dataset more challenging. To evaluate our proposed model for robustness to rotations, we synthetically generated severely rotated coin images. In the presence of rotation differences between coins, our model outperforms the conventional bag-of-visual-words model as well as recently proposed angles histograms of pair-wise identical visual words model.

***Index Terms*—** Bag of visual words (BoVWs), circular tiling, pairwise identical visual words (PIWs), computer vision, support vector machines

1. INTRODUCTION

The BoVWs based image representation has been used in a variety of vision related problems like object category recognition [1, 2], scene classification [3, 4] and large-scale image retrieval [5, 6]. Features are extracted from images and then clustered to form a visual vocabulary. Visual words belonging to the visual vocabulary are then used for image representations. However spatial information of visual words in 2D image plane is not included in the ordinary BoVWs representation. Spatial information of visual words added to BoVWs in a manner that is robust to geometric transformations will lead to improved performance in problems like object category recognition. Therefore for image-based classification of circular objects like ancient coins, a BoVWs model that includes spatial information of visual words performs better than ordinary BoVWs.

Fig. 1: Variations in symbols

Ancient coins are discovered all over the world by Numismatic researchers and hobbyists from archeological sites and fields. Classifying these coins manually is a complex and time consuming task and is usually performed by experts using standard reference books [7]. An image based classification framework that incorporates the knowledge of Numismatics can efficiently assist the task of classifying newly discovered ancient coins.

A *coin type* is an indexing number given by a standard reference book [7] to visually similar ancient coins. Therefore, by ancient coin classification we mean determining the *type* of an ancient coin. Due to the fact that symbols[1] are highly discriminative as compared to the portraits of emperors minted on the obverse side, they can be used as efficient cues for classification. Since a specific symbol may be minted on coins of more than one type, symbol-based coin classification is coarse-grained and can further be refined by other classification cues like legend recognition [8, 9].

Classifying Roman Republican coins using symbol recognition is a non-trivial task because symbols demonstrate significant variations for reasons both ancient and modern. A

[1] In Numismatics, the term "symbol" is used for the by-marks minted along with the main objects on either side of the coins [7]. We use the term "Symbol" for the main object minted on the reverse side.

few examples are shown in Figure 1. For instance, due to the lack of technology, the die engravers were not very precise while minting the coins, therefore a given symbol may be minted differently on different coins of the same type. By-marks minted along the main symbols can cause visual dissimilarities of various magnitudes. As an example, minor variations are caused by the small by-marks minted beneath the belly of "Griffin". However, larger by-marks minted along with the "Curule" cause visual dissimilarity of high magnitude. Another reason for variations in symbols is the absence of relevant parts for visual classification. For instance, the heads of both "She-wolf" and "Dolphin" are absent in their respective images. As coins have possibly been unearthed from fields and archeological sites, these conditions had led to visual aberrations of symbols on coins as can be seen in case of "Triga". Modern reasons of variations among symbols are mostly related to imaging conditions. Circular objects like coins can be imaged under different in-plane orientations as can be seen in case of "Dolphin" and "Curule". Variations in imaging backgrounds and illumination conditions also make our dataset more challenging as it consists of coin images from three different sources which are Kunsthistorisches Museum Wien[2], British Museum London[3] and the coins search engine acsearch.info[4].

Modern coins are accurately minted by utilizing sophisticated technology. Moreover, unlike ancient coins, modern coins do not suffer from corrosion and absence of visually vital parts. Therefore approaches for modern coins [10–12] are not sufficient for ancient coin classification as shown by [13]. In [14], one of the first frameworks developed for ancient coins recognition is presented.Various interest point detectors, local image descriptors and their combinations are evaluated for ancient coins recognition. They use feature matching techniques to assign a test coin image to one of the three classes. Directional histograms calculated at the positions of keypoints are used by [15] to incorporate geometry into SIFT features [16]. Their proposed method outperformed the traditional BoVWs model on images of the obverse sides of Roman imperial coins. [17] proposed to derive visual similarity of ancient coin images by using dense correspondence search amongst them. This visual similarity is then utilized for a coarse-to-fine search to retrieve the most similar coin images in a database. We recently proposed a novel approach by using symbol recognition for image-based ancient coin classification [18]. Symbol recognition for 8 symbols is performed by using the dense SIFT based BoVWs model. Spatial information is added to the BoVWs model using three types of spatial tiling schemes where the rectangular tiling outperformed the other two tilings.

We focus on achieving robustness to coin image rotations as this issue has not been explicitly addressed before. Since the images in our dataset belong to various sources, their backgrounds are different. As a pre-processing step, the backgrounds of coin images are suppressed [19]. Based on our previous work [18], we use symbol recognition for image-based ancient coin classification. Due to degradations in the ancient coins caused by the aforementioned reasons, features are densely extracted from images to capture the underlying structures of symbols. We use circular tilings [18] to add spatial information to the BoVWs model in a rotation invariant way. Furthermore, we model the geometric relationships of identical visual words by using the angles histogram proposed by [20]. However their proposed method is not rotation-invariant. We modify their angles histograms to make them rotation invariant. The modified angles histograms are then combined with circular tilings [18] which not only achieve increased classification rate but also reduce the calculation complexity.

The rest of the paper is organized as follows. Section 2 gives a brief overview of the BoVWs representation of images. In Section 3, we give the details of angles histograms and their combination with the circular tiling. Results are summarized in Section 4 while Section 5 concludes.

[2]http://www.khm.at
[3]http://www.britishmuseum.org
[4]http://www.acsearch.info

2. BAG OF VISUAL WORDS

In the bag of visual words model, a visual vocabulary is built from a set of images. Descriptors like SIFT are extracted from these images which are then clustered using a quantization scheme like the k-means clustering with M number of cluster centers. The size of visual vocabulary $voc = \{v_1, v_2, v_3, \ldots, v_M\}$ is M. A given image is first represented as a set of descriptors

$$I = \{d_1, d_2, d_3, ..., d_N\} \tag{1}$$

where N is the total number of descriptors. A given descriptor d_k is then mapped to a visual word; v_i, using some similarity measures like the Euclidean distance as follows.

$$v\left(d_k\right) = \underset{v \in voc}{\arg\min}\ \text{Dist}\left(v, d_k\right) \tag{2}$$

Where d_k is the k^{th} descriptor in the image, $v\left(d_k\right)$ is the visual word assigned to this descriptor based on the distance $\text{Dist}\left(v, d_k\right)$. In dense sampling based BoVWs, descriptors are collected from images using a grid with a specific pixels stride. Each descriptor is then assigned to a visual word from the vocabulary. An image is represented by the histogram of visual words where the number of bins of this histogram is equal to the size of the visual vocabulary M. The value of the bin b_i of this histogram gives the number of occurrences of a visual word v_i in an image.

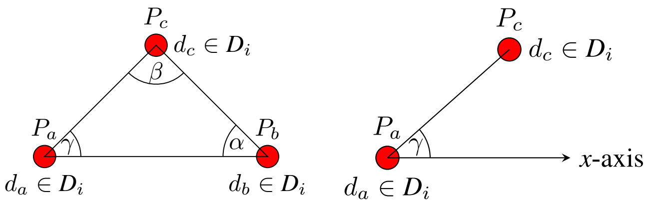

Fig. 2: 3-PIWs and 2-PIWs

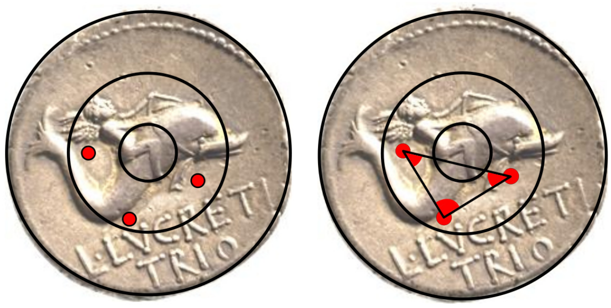

Fig. 3: Simple circular tiling and PIWAHCR

3. PAIR-WISE IDENTICAL WORDS ANGLES HISTOGRAM (PIWAH)

Explicitly modeling similar image cues provides valuable discriminative information about image contents [21]. Recently Khan et al. [20] proposed to use the explicit spatial relationship among identical visual words. Their approach achieved competitive results against the famous spatial pyramid matching model (SPM) [3] on four benchmark datasets. They represent images using normalized angles histograms which are built from angles between two pairwise identical visual words **2-PIWs** as shown in Figure 2. This model is not rotation-invariant as these angles are calculated with respect to x-axis. In this section we extend the model presented by Khan et al. [20] to achieve rotation invariance by utilizing three pairwise identical visual words **3-PIWs** as shown in Figure 2.

Let D_i be the set of all descriptors mapped to a visual word v_i, then the i^{th} bin of the histogram of visual words b_i, is the cardinality of the set D_i.

$$b_i = Card\left(D_i\right) \text{ where } D_i = \{d_k, k \in [1, \dots, N] \mid v\left(d_k\right) = v_i\} \quad (3)$$

From set D_i, all the distinct pairs of three descriptors are considered to calculate angles between the spatial positions of the descriptors as shown in Figure 2. The spatial position of a descriptor is given by its position on the dense sampling grid. The set of all 3-PIWs related to a visual word v_i is defined as:

$$\textit{3-PIW}_i = \left\{(P_a, P_b, P_c) \mid (d_a, d_b, d_c) \in D_i^3, d_a \neq d_b \neq d_c\right\} \quad (4)$$

where P_a, P_b and P_c are the spatial positions of the descriptors d_a, d_b and d_c respectively. The value of the i^{th} bin of the histogram shows the frequency of the visual word v_i. Therefore the cardinality of $\textit{3-PIW}_i$ is ${}^{b_i}C_3$ which is the number of all possible subsets of three distinct elements among the elements of D_i. However, we ignore those pairs of 3-PIWs in which all the three words are collinear. In order to suppress very small angles, we also ignore triangles made by pairs of non-collinear 3-PIWs where one of the three angles is less than $5°$. Therefore for a word v_i, the number of candidate 3-PIWs for angles computation is most likely less than ${}^{b_i}C_3$. The three angles shown in Figure 2 made by 3-PIWs are calculated according to the law of cosines. The angles histogram is built from these angles for which the bins are empirically chosen between $0°$ and $180°$. The angles histogram for a specific word v_i is named as $PIWAH_i$. Khan et al. [20] proposed a 'bin replacement' technique to combine the $PIWAH_i$ from all the visual words. The bin b_i of the histogram of visual words associated with visual word v_i is replaced with $PIWAH_i$ in such a way that the spatial information is added without altering the frequency information of v_i. Finally $PIWAH_i$ of all the visual words are combined to represent a given image.

$$PIWAH = (\psi_1 PIWAH_1, \psi_2 PIWAH_2, \dots, \psi_M PIWAH_M) \\ \text{where} \quad \psi_i = \frac{b_i}{\left\|PIWAH_i\right\|} \quad (5)$$

ψ_i is the normalization coefficient. For a visual vocabulary of size M, if the number of bins in angles histogram is θ, then the size of *PIWAH* is $M\theta$.

Finally, we propose to combine the modified angles histogram with the circular tiling as shown in Figure 3. The 3-PIWs are now limited to the circular tilings only. Doing so not only extracts more information from the circular tilings but also reduces the number of unique combinations of 3-PIWs. We concatenate both the histogram of visual words and angles histogram of all the circular tilings as shown in (6). We call the final histogram *PIWAHCR*, which contains the information of 3-PIWs and visual words other than 3-PIWs for a given circular tiling.

$$PIWAHCR = [PIWAH_1, PIWAH_2, \dots, PIWAH_r, \\ CR_1, CR_2, \dots, CR_r] \quad (6)$$

where for r number of circular tilings, $PIWAH_j$ represents the angles histogram and CR_j represents the histogram of visual words for the j^{th} circular tiling. Therefore for a vocabulary size of M, number of circular tilings r and number of bins in angles histogram θ, the length of *PIWAHCR* is $(r \cdot M \cdot \theta) + (r \cdot M)$.

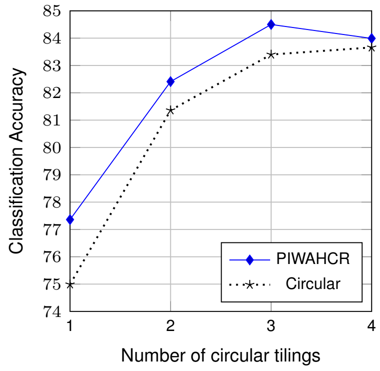

Fig. 4: Performances of PIWAHCR and circular tiling on various number of tilings

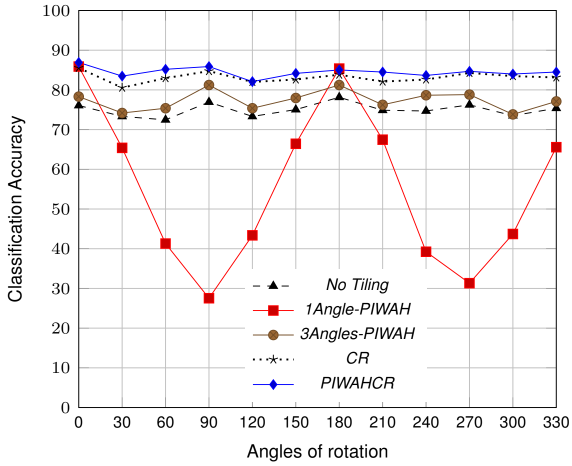

Fig. 5: Performances of various methods on rotated dataset

4. EXPERIMENTS AND RESULTS

For clarity, we define the following names for various settings:

- **1Angle-PIWAH**: Angles histogram made by 2-PIWs as proposed by [20].
- **3Angles-PIWAH**: Angles histogram made by 3-PIWs

Experiments were performed on 2000 images belonging to 15 classes of symbols on Roman Republican coins. For classification we use the one-vs-all approach of SVM. We explicitly compute the feature map for our histograms using Hellinger kernel so that the SVM classifier remains linear in the newly computed feature space. The best value of the regularization parameter C is calculated using n-fold cross validation on training set. The visual vocabulary is also constructed from the training set. The size of visual vocabulary is 100. SIFT features are collected from images using a regular grid of pixel stride 10. We calculate the dominant orientation of each feature to achieve robustness to rotations. Since our dataset lacks challenging rotations, we generate rotated coin images to evaluate our proposed solution for rotation invariance. These images are generated by keeping one image as reference and then rotating this reference image by a step of $30°$. Therefore our test set consists of 12 subsets. We evaluate the number of tilings 'r', in both *PIWAHCR* and simple circular tiling (*CR*). The values of 'r' are empirically chosen as $\{1, 2, 3, 4\}$ where $r = 1$ means no tiling at all. We do not increase the number of tilings as it will result in narrower circular tilings thus reducing the probability of 3-PIWs in a given tiling. The evaluation results for the number of tilings are shown in Figure 4. *PIWAHCR* performs better than *CR* as it contains more spatial information. However, at $r = 4$, the accuracy rate of *PIWAHCR* reduces as the reduced width of tilings results in reduced occurrences of 3-PIWs in all the tilings. Since *CR* histogram is appended to *PIWAH* to form *PIWAHCR* (6), the performance of *PIWAHCR* converges towards that of *CR*.

The results of our experiments for various spatial arrangements are shown in Figure 5 where "no-tiling" refers to the BoVWs model without spatial information. We also evaluated *1Angle-PIWAH* on our synthetically generated test subsets. The performance of *1Angle-PIWAH* follows roughly a cosine curve as the angles are calculated between 2-PIWs with respect to x-axis. Overall the *1Angle-PIWAH* is not robust to severe rotations. We achieve triangular relationship between 3-PIWs with our proposed *3Angles-PIWAH*, which is robust to rotations as can be seen in Figure 5. However, even with increased calculation complexity, the *3Angles-PIWAH* performs inferior to *CR*. Therefore we combine the *3Angles-PIWAH* with *CR* i.e. *PIWAHCR*, to achieve improved performance where the number of circular tilings is 3. To summarize, calculating and combining *3Angles-PIWAH* for each circular tiling achieves global spatial relationships of similar visual words in a rotation-invariant manner. This leads to an increase in classification rate of ancient coins in the presence of severe rotations.

5. CONCLUSION AND FUTURE WORK

A modified BoVWs model for symbol recognition based ancient coin classification is proposed which contains the spatial information in a way that is robust to coin image rotations. This modification is achieved using a combination of circular tiling and angles histogram made from angles between three identical visual words. The proposed model is evaluated on synthetically rotated coin images where it outperformed the traditional BoVWs model, angles histograms made from angles between two identical visual words and circular tiling. In future we plan to reduce the huge number of combinations of identical visual words for angles computation. We also plan to apply our proposed model on highly textured objects.

6. REFERENCES

[1] Gabriella Csurka, Christopher R. Dance, Lixin Fan, Jutta Willamowski, and CÃl'dric Bray, "Visual categorization with bags of keypoints," in *ECCV*, 2004, pp. 1–22.

[2] Jianguo Zhang, Marcin Marszałek, Svetlana Lazebnik, and Cordelia Schmid, "Local features and kernels for classification of texture and object categories: a comprehensive study," *International Journal of Computer Vision*, vol. 73, no. 2, pp. 213–238, 2007.

[3] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: spatial pyramid matching for recognizing natural scene categories," in *CVPR*, 2006, pp. 2169–2178.

[4] Fei-Fei Li and Pietro Perona, "A bayesian hierarchical model for learning natural scene categories," in *CVPR*, 2005, pp. 524–531.

[5] Michal Perdoch, Ondrej Chum, and Jiri Matas, "Efficient representation of local geometry for large scale object retrieval.," in *CVPR*, 2009, pp. 9–16.

[6] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Lost in quantization: Improving particular object retrieval in large scale image databases," in *CVPR*, 2008.

[7] Michael H. Crawford, *Roman Republican Coinage, 2 vols*, Cambridge University Press, 1974.

[8] M. Kampel A. Kavelar, S. Zambanini, "Word detection applied to images of ancient roman coins," in *VSMM*, 2012, pp. 577–580.

[9] O. Arandjelović, "Reading ancient coins: Automatically identifying denarii using obverse legend seeded retrieval," in *ECCV*, 2012, pp. 317–330.

[10] M. Nölle, H. Penz, M. Rubik, K. Mayer, I. Holländer, and R. Granec, "Dagobert - a new coin recognition and sorting system," in *DICTA*, 2003, pp. 329–338.

[11] L. Van Der Maaten and P. Boon, "COIN - O - MATIC: A fast system for reliable coin classification," in *MUSCLE CIS Coin Competition Workshop*, 2006, pp. 07–18.

[12] Reinhold Huber, Herbert Ramoser, Konrad Mayer, Harald Penz, and Michael Rubik, "Classification of coins using an eigenspace approach," *Pattern Recognition Letters*, vol. 26, no. 1, pp. 61–75, 2005.

[13] M. Zaharieva, M. Kampel, and S. Zambanini, "Image based recognition of ancient coins," in *CAIP*, 2007, pp. 547–554.

[14] M. Kampel and M. Zaharieva, "Recognizing ancient coins based on local features," in *ISVC*. 2008, vol. I of *LNCS*, pp. 11–22, Springer-Verlag.

[15] O. Arandjelović, "Automatic attribution of ancient Roman imperial coins," in *CVPR*, 2010, pp. 1728–1734.

[16] David G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91–110, 2004.

[17] Sebastian. Zambanini and Martin Kampel, "Coarse-to-fine correspondence search for classifying ancient coins," in *2nd ACCV Workshop on e-Heritage*, 2012.

[18] Hafeez Anwar, Sebastian Zambanini, and Martin Kampel, "Supporting ancient coin classification by image-based reverse side symbol recognition," in *CAIP (2)*, 2013, pp. 17–25.

[19] Sebastian Zambanini and Martin Kampel, "Robust automatic segmentation of ancient coins," in *VISAPP*, 2009, pp. 273–276.

[20] Rahat Khan, Cecile Barat, Damien Muselet, and CHristophe Ducottet, "Spatial orientation of visual word pairs to improve bag-of-visual-words model," in *BMVC*, 2012, pp. 1–11.

[21] Thomas Deselaers and Vittorio Ferrari, "Global and efficient self-similarity for object classification and detection.," in *CVPR*, 2010, pp. 1633–1640.