Image Anaysis and Deep Learning for Aiding Professional Coin Grading

Xingyu Pan, Laure Tougne

Image Analysis and Deep Learning for Aiding Professional Coin Grading

Xingyu Pan, Laure Tougne
Univ Lyon, CNRS, Lyon 2, LIRIS, F-69676, Lyon, France

ABSTRACT

Coin grading means evaluating coins' physical condition. In this paper, we proposed a process for coin grading by quantification of "unexpected elements" such as scratches and dirty marks. We detect respectively significant and tiny "unexpected elements" with the help of handcrafted filters and Deep Learning techniques. The result of our process, which is close to the manual expert one, is considered as a useful help for numismatists.

Keywords: detection, Deep Learning, coins, grading

1. INTRODUCTION

1.1 Numismatic context

"Hobby of Kings" in the past, coin collecting has become more and more popular with an increasing market. The coin's physical condition is one of the most important factors in determining its market value. A well-preserved example can be worth many times more than a worn one. In this context, standards and scales such as the Sheldon scale [1] have been created to measure the coin's condition by attributing a short note or a numeral score, called **grade**. Measuring a coin's grade based on several indicators of traces of wear (remained relief, scratches, shock marks, dirty marks, surface oxidation, *etc.*) is called coin **grading**. However, even under grading standards, the grade given to a certain coin depends on the person who grades it. Even experienced graders may grade the same coin differently. Furthermore, human investigation, shipment and insurance make coin grading relatively costly and inefficient. The motivation of our study is to make coin grading less costly, more efficient and more objective by using image analysis techniques on high quality photos.

1.2 Previous work

With the development of computer vision and artificial intelligence, image-based coin type identification approaches have been investigated for the recent decade [2, 3, 4, 5]. However, image-based coin grading remains an issue to explore. In the early 1990s, professional grading companies filed patents about a computerized grading system [6, 7]. The system integrates robotics, basic image processing techniques and an online image database. However, this system seems too complicate to apply and to generalize since it requires different images of the coin under various lighting conditions; it includes multiple semi-automatic steps and a large set of "expert rules". Experts also argued that the esthetic aspect of a coin cannot be evaluated by computer [8]. Over twenty years later, those grading companies still rely on "manual" grading. Recently, the smartphone application and online tool Photograde*TM* provides one photo example per grade per US coin type, and users can grade roughly their US coins by searching manually the most identical photo example. Another attempt of image-based coin grading study was conducted by Basset, *et al.* [9]. Assuming that worn US Lincoln cent coins are globally darker than well-preserved ones, they tried to use color histogram to measure coins' grade. However, this assumption is not always true since it exists also well-preserved dark coins for certain coin types. Even though coin grading is less studied, the coin's grade is sometimes considered as a factor of noise for coin recognition issues. Current studies [10, 11] show that the in general higher the coin's grade is, the easier it can be correctly recognized. Nevertheless, exception happens on extremely well-preserved ancient coins because more surface details and reflections increase the difficulty of recognition task. Dagobert is one of the earliest image-based coin recognition systems based on edge features [2]. To make its performance robust, authors used the notions of abrasion and dirty energies to describe respectively "expected" and "unexpected" edges of the query coin compared to a template coin. Such idea of detecting "expected" and "unexpected" elements inspired us for coin grading. Since coin grading is a very precise task, high quality and standard coin photos are indispensable for this study. Moreover, we are focused on quite well-preserved modern coins according to available data provided by a numismatic company.

2. PROPOSED METHOD

We use terms “expected elements” and “unexpected elements” to describe respectively elements belonging to a perfect coin, such as patterns in relief, and those belonging to traces of wear, such as scratches, shock marks, dirty marks, *etc.* For well-preserved coins under study, all “expected elements” are supposed to be present. Thus, coin grading is mainly based on quantification of “unexpected elements”. Other factors such as patina (surface oxidation) mentioned in traditional grading systems are not included in this study. Notice that even for experts, it is difficult to detect manually all “unexpected elements” from images. On the one hand, depth information of scratches and shock marks is unclear through a 2D representation; on the other hand, some “unexpected elements” are too tiny and mixed with the texture of the coin surface. Instead of distinguishing nature of “unexpected elements”, we divide them into two categories, illustrated in Fig. 1:

- **Significant “Unexpected elements” (SU):** Large and evident scratches, shock marks or dirty marks that have a different color from the coin surface.
- **Tiny “Unexpected elements” (TU):** Micro scratches mixed with the coin surface but make this zone grainy.

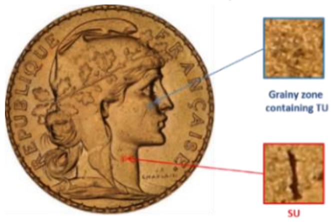

Fig. 1 Examples of SU and grainy zone containing TU on a real coin

At first, the coin is segmented from its photo and identified by using the method of [12]. Date detection is used to improve the identification performance on similar subtypes (varieties) [13, 14]. Then, the coin is registered to a “perfect” reference coin of its identified type. Since coins are segmented, a simple rigid registration is sufficient to find the best rotation. By comparing the coin with the reference coin, we detect respectively SU and grainy zones containing TU. At last, our result, called Grading Guide (GG), quantifies all detected “unexpected elements”.

2.1 Construction of reference coins

In the context of our study, a reference coin should not contain any non-uniformities on its surface expect for contours of relief. However, no real coin is as “perfect” as expected. Assuming that modern coins belonging to the same type have “expected elements” in common but different “unexpected elements”, we construct synthetic reference coins using mean images. Notice that coin relief contours presented in the image, in grayscale, present a sudden change of pixels’ intensities and they are much darker than coin surface background. Let us denote by I_i image of one of the N segmented and registered coins of the same type; we obtain its gradient map I_i^G by using Laplacian operator and its thresholded image I_i^T by using Otsu’s method [15]. Both I_i^G and I_i^T present approximatively relief contours in common. However, we observe that I_i^T contains also some dark regions less illuminated. It is difficult to analyze such regions lack of illumination. Since those less illuminated regions are in common for standard professional coin photos, we also consider them as “expected elements”. Therefore, the binary reference coin mask M is given by

$$M = \frac{1}{N}\sum_{i=1}^{N} I_i^G + \frac{1}{N}\sum_{i=1}^{N} I_i^T \qquad (1)$$

As shown in Fig. 2, white pixels of M are relief contours and dark regions in common. Its black pixels inside of the coin, denoted by $\bar{M}$, are regions supposed to be uniform. Given a registered coin to grade I_k, we detect its “unexpected elements” only in zones of interest ZI_k, given by

$$ZI_k = I_k \oplus M \qquad (2)$$

Fig. 2 Reference coin and zones of interest

2.2 Detection of SU

SU in this context should present a remarkable difference from the coin surface background. For those which present a sudden change in pixel intensity, for example a scratch, we can locate them through image gradients. For those which are darker than the coin surface background, for example a dirty mark, we can detect them by using a threshold. However, neither of the operations mentioned above is perfect. Gradient map cannot detect the entire area of a large dirty mark but

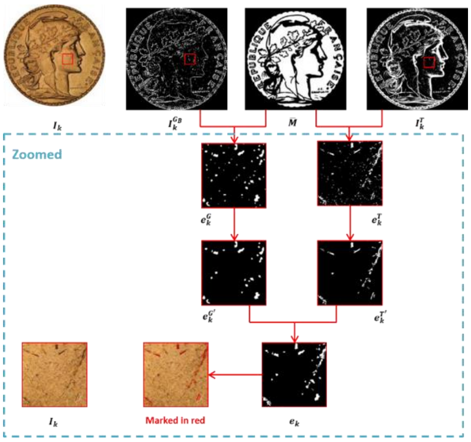

Fig. 3. Process of detection of SU

its external contours, while thresholding techniques cannot capture all slight scratches which are not dark. Thus, we combine both techniques. Let us define e_k^G and e_k^T SU spotted, respectively in the gradient map and in thresholded image. They should be white pixels inside of zones of interest ZI_k, given by

$$e_k^G = I_k^{GB} \cdot \bar{M}, \text{ and } e_k^T = I_k^T \cdot \bar{M} \quad (3)$$

where I_k^{GB} is the binary gradient map of I_k. Notice that both e_k^G and e_k^T contain small noises caused by intrinsically textured coin surface or by tiny "unexpected elements". Regardless of cause, we consider all small connected components as noise and remove them by an experimental threshold. Denoting filtered results by e'^G_k and e'^T_k, the map of SU is obtained by

$$e_k = e'^G_k + e'^T_k \quad (4)$$

Each connected component in e_k is supposed to be a SU. This process is illustrated in the Fig. 3.

2.3 Detection of TU

TU cannot be spotted individually because they are totally mixed with the coin textured surface. It is extremely difficult to analyze this texture since it is a complex combination of texture of metal, texture produced during struck process, mild abrasion and patina. It varies between coins and even between different regions of the same coin. However, according to graders, regions containing TU should be visually more "grainy" than those without TU. Based on this assumption, we can consider that ZI_k of the coin to grade is composed of two parts:

- **Grainy zones:** part of ZI_k visually grainy caused by tiny "unexpected elements";
- **Smooth zones:** part of ZI_k quasi-uniform without tiny "unexpected elements".

To detect grainy zones in ZI_k, we divide ZI_k into small patches in equal size. All patches are denoted by P_k (*cf.* Fig. 4). Using the expert annotation, the goal is then to classify each patch into one of the two classes: "grainy" or "smooth". All grainy patches approximately correspond to the grainy zones containing TU that we want to detect in ZI_k. The main difficulty of this step is that the uniformity of pixels is irrelevant to classification. Only experienced graders can tell which zones are grainy and which ones are smooth. However, as they are unfamiliar with grading coins from photos, it is difficult to precisely annotate data for our study. Despite all obstacles, we have found a professional coin grader to help us annotate "unexpected elements" and grainy zones on hundreds of coins of the same type (20 Francs Coq) on professional coin photos. We name the annotated image database *20FrancsCOQ_Photos*. For patches classification, we extract patches from annotated zones and, put them into grainy class or smooth class according to their original annotated zones. Two observations seem interesting on annotated patches. Firstly, the smaller patches are, the closer are P_k and ZI_k. However, too small patches are too local to represent the graininess that we want to detect. So, according to our experiments, we select a patch size of 64×64 pixels for a segmented coin of 2048×2048 pixels. Secondly, not all annotated smooth patches are more uniform than grainy ones. Sometimes we can hardly find any difference between them (*cf.* Fig. 5).

Fig. 4. Division of zones of interest by patches (64×64 pixels)

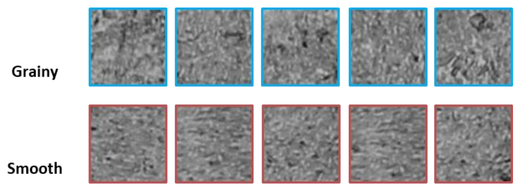

Fig. 5. Examples of indistinct annotated patches in grayscale

As mentioned above, patches classification is almost impossible for human eyes, especially for non-experts. Consequently, traditional machine learning algorithms that require hand-engineering features are difficult to apply. Trying to find a robust method to classify our patches annotated by experts, we use deep convolutional networks. On the one hand, they have shown great performance on various image classification and object recognition issues over traditional algorithms. On the other hand, the network learns the filters applied to original patches. A general structure is composed of convolutional, pooling and fully connected layers. In practice, instead of training an entire CNN with random initializing, fine-tuning techniques are used to adapt a successful pre-trained model to a similar problem. In this study, we chose to use a fine-tuned AlexNet [16] model according our test.

2.4. Grading Guide (GG)

The result of our system, called Grading Guide (GG), is composed of the surface of SU, the number of SU, and the ratio between grainy zones (containing TU) and zones of interest. Notice that GG is not a grade that can be directly used in the market, but it serves as a relatively objective indicator of "unexpected elements" of the coin to help graders attribute a numeral grade more precisely. The numeral grade G_N can be given by

$$G_N = f(\mathrm{GG}, W) \tag{5}$$

where W is the set of experimental weights associated to different elements in GG. This part will be studied in the future with the help of professional graders.

3. EXPERIMENTS AND DISCUSSIONS

In this section, we evaluate the performance of our method on the database 20FrancsCOQ_Photos. It is difficult to validate quantitatively the performance of detection of SU due to lack of properly annotated data for SU. In Fig 6, we can see that large scratches and dirk marks can be correctly spotted by the proposed method. For detecting grainy zones containing TU, we constructed different datasets with various number of patches extracted randomly from all annotated coins. All databases contain N_g grainy patches and N_s smooth patches with $N_g = N_s$. For each dataset, we used two thirds of patches as training set and one third as test set. In both training and test sets data are equally distributed between two classes. As input of AlexNet, three types of patches were tested: original color patches, grayscale patches and normalized grayscale patches (*cf.* Fig. 7). According to the results obtained by cross validation shown in Table 1, classification rates are generally high by using 2000 to 3000 patches but decrease if more patches are used. However, when we use 9000 patches, classification rates increase again but cannot match the results obtained on data with a small size. Another observation is that color patches seem more performant than grayscale ones on data with a small size. However, the more data we used, the closer are results obtained by using different types of patches. It may explain that color or global intensity is irrelevant to the graininess of patches. We also conducted experiments by using the same total patches in different class distribution. From the results shown in Table 2, we can see that the distribution of data influence much less than the size of data. It seems that an increasing size of data will make those two classes less distinct. We examined also our results with the grader who annotated our data. In general, we detect grainy zones with similar sizes and at similar locations to the annotation *(cf.* Fig 8). Interestingly, some wrong classified patches are reannotated as correct after human verification. In fact, according to the grader, his annotation is far from precise for many reasons, which explains why more trained data increase the confusion. In a numismatic view, the objective of our study is not to obtain results close to one "subjective" annotation but to provide a new version of grainy zones that will be accepted by most experts. In other domain of Deep Learning, Alpha Zero has already proved human annotation could be noise [17]. From that point of view, our results are encouraging.

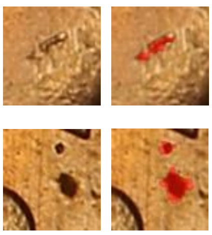

Fig. 6. Examples of detected SU (in red) on zoomed zones of interest

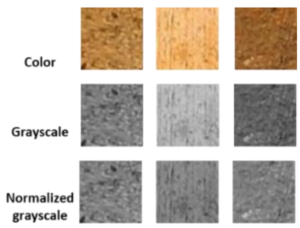

Fig. 7. Examples of tested patches

Fig. 8. Examples of detected grainy zones

Table 1. Results obtained on data in various size

Total	N_g	N_s	Color	Gray	Norm. Gray
2000	1000	1000	85.3%	84.4%	83.3%
3000	1500	1500	86.2%	84.2%	82.8%
5000	2500	2500	73.5%	74.2%	73.3%
9000	4500	4500	76.5%	75.7%	75.6%

Table 2. Results obtained on data in various distribution

Total	N_g	N_s	Color	Gray	Norm. Gray
3000	500	2500	85.0%	85.2%	84.1%
3000	1000	2000	85.3%	84.4%	83.3%
3000	1500	1500	86.2%	84.2%	82.8%

4. CONCLUSION AND FUTURE WORK

In this paper we have presented a comprehensive method to aid professional coin grading. After constructing a synthetic reference coin to compare, we detect significant "unexpected elements" by using image gradients and thresholding techniques, and grainy zones containing tiny "unexpected elements" with the help of Deep Learning techniques. The proposed method is promising according to professional graders. However, the proposed method was only tested on the available small database of professional coin photos. We are going to test it on large databases including more coin types. One of possible extensions is to qualify "expected elements" on worn coins. In addition, scoring patina by color or texture analysis is important to complete the last brick of coin grading. Furthermore, new acquisition approaches will be studied since depth information of "unexpected elements" is more critical than their size. According to professional graders, a deep scratch on the coin surface is much more severe than a shallow one or a dirty mark in a similar size.

REFERENCES

[1] Sheldon, W. H., Paschal, D. I., and Breen, W. H., [Penny whimsy: a revision of early American cents], Quarterman Publications (1976).

[2] Nölle, M., Penz, H., Rubik, M., Mayer, K., Holländer, I., and Granec, R., "Dagobert-a new coin recognition and sorting system," in Proc. 7th Int. Conf. on Digital Image Computing-Techniques and Applications (2003).

[3] Nölle, M., Rubik, M., and Hanbury, A., "Results of the Muscle CIS coin competition 2006," in Proc. Muscle CIS Coin Competition Workshop, 1-5 (2006).

[4] Zaharieva, M., Kampel, M., and Zambanini, S.,"Image based recognition of coins–an overview of the coins project," (2007).

[5] Kavelar, A., Zambanini, S., Kampel, M., Vondrovec, K., and Siegl, K., "The ILAC-project: Supporting ancient coin classification by means of image analysis," in Proc. 24th Int. CIPA Symp., (2013).

[6] Merton, H. A., "Method and system for objectively grading and identifying coins," U.S. Patent No. 4,899,392. U.S. Patent and Trademark Office (1990).

[7] Crain, L. M., "Automatic coin grading system," U.S. Patent No. 5,220,614. U.S. Patent and Trademark Office (1993).

[8] Haperin, J., [How to Grade U.S. Coins] (1990).

[9] Bassett, R., Gallivan, P., Gao, X., Heinen, E., and Sakalaspur, A., "Development of an automated coin grader: a progress report," in Proc. 8th Annual Mid-Atlantic Student Workshop on Programming Languages and Systems, (2002).

[10] Fare, C., and Arandjelović, O., "Ancient roman coin retrieval: a systematic examination of the effects of coin grade," in Proc. European Conf. on Information Retrieval, 410-423 (2017).

[11] Schlag, I., and Arandjelovic, O., "Ancient Roman coin recognition in the wild using deep learning based recognition of artistically depicted face profiles," in Proc. ICCV Workshop, 2898-2906 (2017).

[12] Pan, X., Puritat, K., and Tougne, L., "A new coin segmentation and graph-based identification method for numismatic application. in Proc. Int. Symp. on Visual Computing, 185-195 (2014).

[13] Otsu, N., "A threshold selection method from gray-level histograms," IEEE Transactions on Systems, Man, and Cybernetics, 9(1), 62-66 (1979).

[14] Pan, X., and Tougne, L., "Topology-based character recognition method for coin date detection," in Proc. IEEE Int.Conf. on Image Analysis and Processing, 2, (2016).

[15] Pan, X., and Tougne, L., "A new Database of Digits extracted from coins with hard-to-segment Foreground for Optical character recognition evaluation." Frontiers in ICT, 4, 9 (2017).

[16] Krizhevsky, A., Sutskever, I., and Hinton, G. E., "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, 1097-1105 (2012).

[17] Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., ... and Chen, Y., "Mastering the game of go without human knowledge." Nature, 550(7676), 354 (2017).