Coin Wear Estimation and Automatic Coin Grading

Parmida Atighehchian

A Thesis

in

The Department

of

Computer Science & Software Engineering

Presented in Partial Fulfillment of the Requirements

for the Degree of

Master of Computer Science (Computer Science & Software Engineering) at

Concordia University

Montréal, Québec, Canada

April 2017

© Parmida Atighehchian, 2017

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Parmida Atighehchian**

Entitled: **Coin Wear Estimation and Automatic Coin Grading**

and submitted in partial fulfillment of the requirements for the degree of

Master of Computer Science (Computer Science & Software Engineering)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the Final Examining Committee:

______________________________ Chair
Dr. Charalambos Poullis

______________________________ Examiner
Dr. Tien Bui

______________________________ Examiner
Dr. Brigitte Jaumard

______________________________ Supervisor
Dr. Ching Y. Suen

Approved by ______________________________
Sudhir Mudur, Chair
Department of Computer Science & Software Engineering

__________ 2017 ______________________________
Amir Asif, Dean
Faculty of Engineering and Computer Science

Abstract

Coin Wear Estimation and Automatic Coin Grading

Parmida Atighehchian

In numismatic studies, coin grading is referred to as the set of detailed experiments on a coin in order to estimate its quality, which is the most important factor to estimate the coin's value. Usually, the task is done by three expert numismatists to minimize personal biases. Each numismatist tests the coin's wear, coloration, and toning under different lighting conditions. Coin grading is a sensitive task to be done by humans. There are different parameters that can define the coin's value, however, dependent on the numismatist expert conducting the test, some parameters are neglected and some are given a heavier weight, which makes the procedure very subjective. A computer-aided algorithm for coin grading is considered an asset to help conduct more objective coin grading experiments.

We propose a coin wear estimation algorithm, which is fully based on features extracted from the digital images of coins. Apart from coin grading, the proposed algorithm is useful to find and dismiss the heavily worn out currency from the market. As online trading is getting more and more popular among coin collectors, it has become easier for individuals to sell a low-quality coin instead of a high-quality one or foist fake copies instead of real coins. This study is concentrated on the feasibility of having a computer-aided program to conduct coin grading. The required specifications for the dataset are fully investigated and the final dataset is collected after lots of experiments. In our proposed method, SIFT key points are used to distinguish the amount of wear on the coins. These key points are known for their high accuracy in shape detection problems. Our approach in using these descriptors to estimate the amount of wear on the coins attains a high accuracy of 93%.

Acknowledgments

I would like to take this opportunity and send my gratitude to those who helped me in preparation of this thesis.

First and foremost, my deepest respect and gratitude goes to my supervisor, Dr. Ching Y. Suen. He inspired me to be independent, positive and always aiming the best. During my studies, He has been always a great support for my growth and taught me to be innovative. I hope I have delivered my commitments as a qualified master student and could use his lessons for the rest of my life.

I would also like to thank my dearest friends, Pedro Maroun Eid, Mathieu Germain and Ehsan Amiri who have been great supports with their constructive comments during my master studies and without whom I could never go this far. A special thanks to Pedro, because he taught me never stop fighting in my life.

I would like to thank Nicola Nible and Marleah Blom who helped me with data collection and proof reading of some chapters. I am so grateful that I had the privilege to work with them. I would like to thank the examination committee for taking the time to read and help me make this documentation even better with their valuable comments.

I would like to thank my uncle, Dr. M. R. Ghanoonparvar, and my aunt, Dian Wilcox who were always supportive and ready to help me in this period. I really appreciate their emotional support which was exactly what I needed when I was far from my family.

Last but not least, my greatest appreciation goes to my parents and my little sister, Sourena, who always taught me to be strong and think big of myself. The dearest people in my life who even though were not physically present, I did not feel their empty spot in my life even for a second.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 What is Coin Grading

Coins have been one of the fundamental elements of human civil life since ancient times. In the recent century, coin collecting and trading has evolved as a big business and many ”Numismatic Institutions” have started to work on this business. Similar to other goods, there should be a valid scaling system to evaluate the quality of the coins for sale. The word ”Numismatics” is referred to as any study related to coins[2], and the main responsibilities of numismatic institutions are to evaluate the quality of the coins. The process of evaluating the coin’s relative quality and condition is named as coin grading in numismatic studies[2]. The numismatic institutions act as third-parties in the coin collecting market. These institutions grade the coins, independent of both the buyer and the seller’s influence. The most well-known numismatic institutions are the American Numismatic Association Certification Service (ANACS), the Professional Coin Grading Service (PCGS), and the Numismatic Guaranty Corporation (NGC).

Early coin grading was as simple as assigning the coins to one of the defined three categories:

- *Good*: The details on the coin are still recognizable but the texture is worn out.
- *Fine*: Mint luster can be seen to some extent.
- *Uncirculated*: The features are very sharp and the luster band is fully recognizable.

Today coin collecting has evolved as an independent market and the above simple three level scaling system can not serve the demand of this big market anymore. Today's grading scales are much more complicated. There are several grading standards that are commonly used such as the European grading system and the American Numismatic Association (ANA) grading standards. These grading standards find their roots in the well-known "Sheldon grading scale" in which a scale of 1 to 70 is given to a coin based on its quality. Most of today's grading standards have more or less the same body structure as "Sheldon grading scale", which makes the standard universal and legitimate. A more detailed description of the main structure of the commonly used grading systems is given in Appendix A.

1.1.1 Principal Components of Coin Grading

There exists more than sixty grading standards and increasingly more coin specialists use these standards as their guide to grade. Grading is a procedure, which is mainly based on experience. There are so many different elements that may be considered when grading a coin. All these elements have more or less tangible definitions which give an overall sense when studying the coin. The most common elements are listed below:[2]

- *Eye appeal*: This is considered one of the important factors in deciding the grade of a coin. It is affected by several different parameters, such as the amount of luster, coloration, and toning. Luster is referred to as the reaction of light when it hits the mint-state coin's die erosion lines. It usually creates an appealing reflection of the light when moving the coin under the light and is a positive element in grading especially for the mint-state series. Coloration and toning are referred to as the change of color in different coins based on the amount of reactive metals used for them (mostly copper and silver). It might cause a change in red color of copper coins, turning it more brown or it can create a beautiful hue on the surface of a silver coin. The first effect is not appealing to the eyes and has a negative impact on the final grade whereas the latter has a positive impact in grading. Coloration and toning are considered especially when dealing with ancient coins' grading.

- *Contact and Detracting marks*: These marks are caused by mishandling or cleaning the coins with improper materials. These marks appear on coins on the market very soon after their release; however, a collector can prevent these marks to appear on his collection to some extent. The severity of these marks has an important role in assigning the coin to one of the described grading levels in Appendix A. Scratches and wear effects will fall into this category as well. These marks have a greater weight when deciding among the Fine category levels (see Appendix A).

- *Strike marks*: In the process of striking a coin, the designed die is used several times until it is changed. Therefore, even the mint-state coins do not have the same grade as the quality of the die and the strike is important. When the coin has a sharper profile and a greater deal of details, the higher the grade is. Die flaws are also considered as a negative factor in deciding among the mint-state categories. Depending on the severity of the marks, the coin's grade varies.

- *Obverse and Reverse grade*: Although both the obverse and the reverse of a coin have important roles in defining the final grade, the attention is mostly directed to the front. Generally, the obverse defines the category of the coin's grade; however, when there is a tie between two coins based on their obverse, the reverse plays its role. A rule of thumb is to consider two times higher importance for the obverse, although it can be changed based on the specialist's opinion [2][3].

An important fact to be considered is that not all of these parameters are always taken into account when a professional is grading a coin and that is where experience comes in. An experienced coin specialist goes for the most important elements to consider for each type of coin. Even if we want to be precise and quantitative, it is impossible based on the given factors as the first factor is very subjective and experience oriented. Therefore, the whole procedure of grading a coin, even when graded based on three different specialists' opinions, could be questionable in some cases. This brings us to a need for a computer-aided algorithm which can facilitate the work and gives a higher confidence to the final stated grade.

1.2 Problem Definition and Motivation

Coin grading as explained in sections 1.1 is only used for the coin collecting market. However, a deeper look to the field reveals other important applications which brings us to the point to feel the urge of having a more stable system for grading. Apart from being an essential component in coin collecting market, coin grading can be used in:

(1) Detecting counterfeit coins

(2) accelerating the process of rejecting damaged coins at banks

(3) improving numismatic institutions grading systems

(4) Offering a quality measurement tool for online coin buyers

Coin collectors are not the only group who are facing the problem of detecting and dealing with counterfeit coins. In some countries, counterfeit coins are made and fed into the system as part of the daily purchases. China and Denmark have been facing lots of problems in their banking and trading system. While Canadian coins might seem to be cheap, they are not far from being faked. In a report by Royal Canadian Mounted Police in 2006, a company making counterfeit Toonies and Loonies has been spotted near Montreal. Nowadays, counterfeiting is not a rudimentary process any more. In many cases a counterfeit coin can not be distinguished by its weight, color, dimension or any other physical measurements, which brings us to precise study of the coin's pattern as a solution for modern counterfeit detection. Although the planchet (the round metal disk which is getting struck) is available to everyone, the die itself is where the major differences between a genuine and a counterfeit coin could be found. Usually the counterfeit coins tend to be worn out faster than the genuine ones and die mistakes are a lot more common in them. Figure 1.1 is an example of such case. The coin on the left is the genuine while the coin on the right is a fake one. This is where coin grading is beneficial.

(a) A genuine coin with sharp and clear edges

(b) A fake coin with a poor quality design

Figure 1.1: Comparison between a genuine and a fake Danish coin

Coin grading can also be beneficial in banks to spot and dismiss heavily worn out coins off the market. The main problem in today's coin grading is that it is dependent on human decisions and experience. As mentioned in the previous section, there are lots of elements to be considered when grading a coin and it is only based on the grader's opinion to choose which ones to consider. Coin grading is a sensitive task to be done by humans, since most of the time it involves a huge drop in the coin's price when it is graded as a lower quality coin. Although having three professional graders will help to balance personal judgments on the final grade, an automatic grading system will definitely be a better option to have a more objective grading system. A computer-aided coin grading system is also useful for collectors. As online trading is growing, a system that provides a reliable grade based on the digital image provided on the coin trading website is crucial. Online trading is where most false information could be provided and the computer-aided coin grading can prevent fraud within the coin collection business.

Automatic coin grading is not an easy task, especially when the goal is to only use digital images. The challenges involved in this field of study can be listed as below:

- In coin grading, eye-appeal has an important role, which is a really hard factor to incorporate in image studies. However, except for the use in the numismatic institutions as a grading system, generally a coin wear quality estimator needs to be focused on the overall wear, and detracting marks to cover the applications listed above.

- Finding equivalent image processing measurements for common grading factors and a proper scale to find out which one has a greater weight in different situations is not a trivial task.

- The algorithm needs to be general and not dependent on a specific design. This reveals another big challenge that the design differences shall be managed differently from marks and scratches. Using image processing techniques, this matter is the most challenging concept. In other words, to have an algorithm which is only triggered by the wear changes.

- The algorithm should have enough precision to at least successfully distinguish among the different existing general labels of uncirculated, almost uncirculated, extremely fine, very fine, and fine.

- Dealing with images has the disadvantage of having illumination effects in the case of coin studies. Illumination variance affects the accuracy to a high extent by false detection of scratches as edges or false classification of a worn out edge as a sharp one and vice versa due to uneven lighting or changes in the material of the different parts of the coin as in the case of Toonies.

1.3 Literature Review

Coin grading is a fresh topic in coin studies literature and not many studies could be found to be directly focused on the topic. However, the other related studies could be very useful to be reimplemented in a compatible fashion to this study or to inspire us to chose a suitable approach. Among the related studies, ”coin recognition” and ”counterfeit detection” are the most relevant.

In [4] and [5], a summary of the existing image processing methods for coin recognition is provided. Reference [4] is focused on the different methods proposed for modern and ancient coins. There are two main segmentation steps proposed for coin studies. Using the Hough transform algorithm, which is proposed in [6] is an exact solution given for perfectly round shaped coins. However, the edge detection methods followed by morphological processing, which was used in [7], is applicable to any shape, which makes it more suitable for ancient coins studies. The three main challenges included in every coin study are about how to handle illumination, scale, and rotation changes. Different coins have different diameters, which although this is an important feature used in most mechanical forensic detection systems, it cannot be considered as a stable feature in image studies of coins unless the camera is set up properly and the setup is not changed through the whole process of image capturing. Either by a fixed setup or direct input of the diameter, this physical measurement together with the thickness of the coin have been used in so many studies to uplift the recognition rate [8][9][6][10]. Other studies chose to have a scale correction step prior to their algorithm in the preprocessing step. Some studies have a step to correct the rotation changes using correlation and registration techniques [10]. Using rotation invariant features is an intelligent approach, which was used in many recent studies [8][9][11][12]. The features are either by nature rotation invariant [11][12] or extracted in a concentric ring structure, which is the trick to make the features invariant to rotation[8][9][13][14][15].

The differences between ancient coins and modern coins affect all the processing steps from coin extraction to feature selection. The main challenge is the arbitrary shape of ancient coins, which makes the Hough transform useless for extraction step. The texture and wear of ancient coins are severely damaged and they are either hammered or cast whereas modern coins are minted. Generally speaking, mint is a much more precise procedure, which leads coins from the same type to have stable common features and makes the coin recognition algorithms to be more straight forward and a lot easier to be generalized. However, it is completely a different case in ancient coin recognition. Based on the mentioned textural and casting procedure differences, the intraclass variance in this category of coins is much larger. This is an indication of larger texture variation in ancient coins which differs the problem-solving method from the one for modern coins [4][11]. There are many features used in coins studies including, but not limited to, edge information-based features, gradient features, texture features and shape context-based features. Below is a summary of the most important coin recognition methods used in the literature.

1.3.1 Edge Information and Contour-based Algorithms

Perhaps the implementation of Dagobert was a peak in the studies concentrating on numismatics and coin recognition. This automatic coin recognition system was developed to have speed together with high accuracy for the recognition of coins from 30 different classes [8]. The coin edge image is compared to a pre-selected list of master coins and the largest correlations are selected to identify the coin's type probable categories. The final category is selected using three rotation invariant features, edge-angle distribution and edge-distance distribution, which are compared to the center and are calculated on the concentric rings, and counting the occurrence of different rotation-invariant patterns on the circles centered at edge pixels. The diameter and the thickness are also calculated as the physical measurements which help the procedure of pre-selection of the master coins. The final reported accuracy on a testing set of 12949 coin images is 99.24% .

Maaten et al. in [9] developed a fully automatic coin recognition system that was tested on the famous MUSCLE dataset. The training set used in this study contains images from 20000 coins, which fall into 692 different classes. The testing set contains images for 5000 coins. Both the training and the testing sets contain images from obverse and reverse of each coin. The main challenge for their study, apart from speed, was to have a correct rejection mechanism. For this reason the testing set contains coins that are not categorized in any of the training set classes. Their algorithm follows the routine procedure of coin recognition, which includes segmentation, feature extraction, classification, and validation. In order to have the speed, segmentation includes thresholding, edge detection, and morphological processing. The extracted statistical edge-based features fall into three categories of edge distance distributions, edge angle distribution, and edge angle-distance distribution. The first and the third features are extracted in circular concentric divisions of the coin, which certifies the rotation invariance. In order for the second feature to be rotation invariant, the magnitude of the Fourier transform of the histograms extracted for each pie shape division on the coin is calculated. The 3-nearest neighbors classifier is used to classify the obverse and the reverse of the coins separately. The final decision is made by a voting system based on the resulting label for each side of the coin. The accuracy gained for this experiment is 72% and misclassification rate is reported for only 2%.

1.3.2 Gradient-based Algorithms

Reference [6] is concentrated on the use of gradient angle information for coin recognition. Despite the dataset used in [8] and [9], in their dataset the coin and the background color are identical. This makes the coin extraction step a bit more challenging and urges the use of the Hough transform in their study. In the feature extraction step, Reisert et al. [6] indicate the disadvantage of using the edge magnitude information and base their features on the edge orientation information. They claim that using the magnitude has the disadvantage of not being consistent with respect to illumination changes. The edge extraction methods are also error prone towards low-quality images. The Fourier transform of the edge direction is then used as the feature for comparison to speed up the process. As in most of other references, the k-nearest neighbors algorithm is used for the classification. The final recognition rate for a benchmark of 10000 coins is 97.24%. However, the physical measurements such as thickness and radius are still used in the procedure to have more confidence on the rejection criteria.

1.3.3 Eigenspace Approach

In [10], Huber et al. use the eigenspace to classify the coins. Rotation invariance is gained by cross-correlation of each image to the reference image and finding the rotation angle. However, physical measurements such as diameter and the thickness are involved in categorizing the coins in the first place and selection of the references. The eigenspaces are selected based on the referred physical measurements, meaning each eigenspace defines a portion of thickness and diameter range. Using eigenspace eliminates the illumination undesirable changes to an acceptable extent for coin recognition. The final decision is made using Bayesian fusion classifier and using both the obverse and the reverse of the coins. The accuracy reported for their experiment is 93.23% and a final rate of 6.77% for false decision.

1.3.4 Texture-based Recognition

Zaharieva et al. [11] use an adaptive thresholding algorithm for segmentation of ancient coins. Both the shape context matching and SIFT features are used and the results are compared. The final result using SIFT shows a 93.93% classification rate for a dataset of 350 coins containing 3 different types of coins. No recognition rate is reported for shape context matching method due to the need for qualification of the proposed algorithm using a much larger dataset. This shows the reliability of SIFT features even using a limited number of images. Another study in [16] shows the concentric dense feature extraction, using the LIDRIC features (Local Image Descriptor Robust to Illumination Changes). The LIDRIC features are known to be reliable against illumination changes. The scale correction is handled with a coin segmentation step prior to that and in order to have more reliable results, the geometric displacement of the recognized key points should not be more than a threshold. The results show an improvement comparing to the other works using SIFT and bag of visual words. Other available literature focuses on extracting different textural features accompanying with the geometrical information. Reference [13] uses local binary pattern features, Gray level co-occurrence matrix and a combination of both in a concentric ring and fan shape structure. Shen et al. [14] compare the result of coin classification using different texture features. In their experiments statistics of the local binary pattern and a set of Gabor wavelets of concentric ring structure sections are used as features. The wavelet coefficients are recognized as a common texture descriptor feature [15][17]. The aim is to break down the image with respect to frequency and direction. The chosen direction(s) for wavelet decomposition and the number of frequency levels to break down the image is decided based on the highest result in accuracy. To deal with the rotation changes concentric ring structure is used in [15]. The wavelet decomposition has generally shown high accuracy rate for texture recognition apart from the type of images in the dataset. Using the Gabor feature coefficients extracted in a concentric ring structure and comparing the extracted feature vector is another proposed method for coin recognition [18].

1.3.5 Character-based Recognition

In 2013, Kavelar et al. [19] proposed a pipeline for coin legend recognition using SIFT features. In their study, they claimed that OCR techniques fail to recognize the characters when the text and the background share an identical color. Therefore, in their proposed coin legend recognition pipeline, they encourage the use of SIFT features in the presence of the SVMs as the classifier. In order to deal with the illumination changes, the features for the opposite lighting angle of the same character are mapped to each other. In their study, the training set consists of 900 images with the resolution of 100×100 and 90 test images. Their proposed pipeline reaches to the accuracy rate of 75.6%. Using the proposed pipeline, Kavelar et al. [12] have used SIFT descriptors to read the Roman Republican coins' legends and classify the coins accordingly. Characters on the Romanian coins are categorized in 18 classes. As described in their proposed method for character recognition, SIFT features are used. Although the features are rotation invariant, the experiments have shown a lift in the accuracy rate when the rotation is fixed. The algorithm is tested on different datasets and the recognition rates changes between 29% to 67%. Despite the fact that the recognition rates are not considered desirable, they claim that the algorithm has the advantage that is not limited to the text recognition on the flat surfaces and bimodal images. Anwar et al. [20] give an interesting approach for classification of ancient coins based on bag of visual words and SIFT features. However, as for coin classification, spatial information makes noticeable changes in the final decision, Anwar et al. introduce merging the spatial information with SIFT key points using different geometrical coin partitioning as rectangular tiling, log-polar tiling, and circular tiling. Based on their experiments adding the spatial information will outperform the usage of a sole bag of visual words. For a dataset of 3900 coins with 550 different types, an accuracy rate of 95% is reported for their study. An extension of this study is given in [21], in which a larger dataset is tested. The different parameters such as the size of the pixel stride, number of features per image to construct the visual word and size of the visual vocabulary are tested. The SVM one-vs-all is used to train the dataset.

1.3.6 Recognition using Neural Network

Neural network is a powerful algorithm which has a long history in coin recognition. In combination with other features to increase the precision of the recognition [22][23][24], it has shown a promising path for coin recognition. Given a large dataset, and using back propagation system a high accuracy of 96.3% is reported to recognize between 1 TL and 2 EURO coins [25]. Using similar approach and extract a feature vector from the pattern averaging step, back propagation neural network gives an accuracy of 97.74% to recognize four different indian coins [23].

1.3.7 Counterfeit Detection

Counterfeit detection studies conducted at Center of Pattern Recognition and Machine Intelligence at Concordia (CENPARMI) started in 2015 and have made a great contribution to the field of numismatic studies by publicizing the pattern recognition techniques used in coin recognition in [3]. Reference [26] is concentrated on the counterfeit detection using different statistical contour-based features on the legend of the coin such as relative distances of the characters to the centroid, stroke width, and the distance to the centroid for each character to recognize the counterfeit. In [27], Khazaee et al. propose a novel approach for counterfeit detection using the energy from the decomposed signals of hight-map images. Their proposed method using an SGD classifier gives a true positive rate of 97.8% with the false positive rate as low as 7.8%.

1.4 Outline of This Study

In this thesis, the feasibility of having a completely automatic coin grading system is investigated. No physical measurements have been used for the study and decisions have been made solely using the digital images. The coin grading is a different topic from coin recognition; however, the features used in the coin recognition literature can be reused in a compatible manner with this study. It is worth mentioning that color variation factors are neglected in this study so that the outcome can be as general as possible to cover all the proposed applications. Hence, the concentration is on the texture differences for different levels of coin degradation. The texture differences can vary from wear to appearance of scratches and bumps on the surface of the coins.

Chapter 2 is concentrated on the dataset preparation. The image acquisition techniques have been investigated and a proper setup for coin grading data collection is proposed.

Chapter 3 proposes the preprocessing pipeline for this thesis. The coin extraction and scale correction have been included in this chapter, also rotation and illumination correction approaches have been proposed and used in for the verification of the final results.

Chapter 4 is focused on the visual differences on the coins with different degradation levels. Although the visual differences are noticeable, the conclusion indicated in the last chapter shows the limitation of utilizing only visual features for coin grading.

Chapter 5 talks about the feasibility of having an automatic grading system showing the distribution of coins' texture with respect to their principal components. The SIFT features are introduced as a perfect choice for this type of texture study and the reasoning behind this selection is provided.

The whole algorithm is thoroughly tested from different aspects and the validity of the algorithm despite the limitations for the project is proved in chapter 6.

The contributions of this thesis and the possible future improvements are listed in chapter 7.

Through out the thesis, some numismatic terminologies are frequently used. A complete list of these words together with their definitions are given in Appendix A. A summarized Sheldon grading scale table compatible to the quality classes used in this thesis is also provided at the end.

Chapter 2

Dataset Collection

2.1 Dataset Preparation

Numismatic studies involve collecting a proper dataset as the first step. The dataset should be large in number to contain enough data samples in each output class. In studies based on image processing and pattern recognition techniques, this addresses the most important part of the problem which needs to be solved, since the validity of the final results is highly dependent on the variety of the dataset and the numbering of the available samples in each class. Computer-aided coin grading is a new subject in the field of numismatic studies, and there exists no such dataset to serve the purpose of this study. Therefore one needs to go through the procedure of data collection. Collection of coins with different texture qualities is highly challenging compared to other coin studies such as coin recognition and fake coin detection, because the output classes for the latter named studies are completely known whereas the output classes for coin grading even if studied by the coin specialists are very subjective and could be changed even based on the variety of different wear qualities within the dataset. The involved challenges in data collection specific to this study are:

(1) **Class frequency:** The optimum dataset should contain coins with different qualities. It is almost impossible to create a dataset from ancient coins, since the good quality and mint-state samples are impossible to be found in large scale. Therefore, the dataset should be created from the current currency.

Figure 2.1: Examples of two different patterns in Toonies

(2) **Within-class variance:** The changes in the designs of the obverse and reverse in different coins is much more impressive (from the pattern recognition point of view) than the changes in the field wear. This causes a negative impact on the final result of the study since the main parameter's (the field wear) changes are much less than other changes in the coins. As a result, the study should be focused on one certain kind of coin. Even in one specific kind of currency, different designs are found from year to year (see Figure 2.1). Focusing on Toonies, for example, the year's spot and the queen's shape are changing from year to year on the obverse of the coin (see Figure 2.2). Investigating into a specific year revealed that even in a certain year small variations in the obverse/reverse design pattern may be found (see Figure 2.3). To decrease the within-class variations, these changes should be kept as low as possible however we cannot limit the changes too much, since collecting a dataset

Figure 2.2: Year's spot displacement and small pattern variations in Toonies

of coins with different qualities is almost impossible considering the limitations of time and resources. Therefore, the data collection is focused on the years which have the most similar obverse design patterns. Differences such as displacement of the year on the coin, having maple leaf marks on some coins, and the small changes in the design pattern of same year coins are tolerable.

Figure 2.3: Small pattern differences in the same year stamped coins

(3) **Shortage in resources:** The coin stores only sell the currency in sealed packages. The high quality coins were collected from coin stores and coin dealers. However, lower quality coins are very hard to find. The banks provide coins in rolls which contain different qualities and texture patterns. The same problem exists with regard to grocery stores and shops. The low quality coins were then collected from the banks going through the exhausting process of buying rolls of coins, taking out the coins with the desired quality and year and returning the reminder to the banks. This process was done over and over again to prepare a valid dataset for this study. In this selection process, only coins from 2009, 2011, 2013 and 2015 (because of the acceptable range of variance and similarity existing within the coins from these years) and with a noticeable amount of wear (since the high quality coins were already acquired from the coin stores) were considered as interesting candidates.

Due to the above challenges in data collection, the procedure was extremely frustrating and time consuming. We managed to prepare 129 coin samples containing different qualities. The coins were then labeled by a coin specialist based on their qualities according to the ”Sheldon scale”. These labels are used in classification, which is discussed in chapter 5.

2.2 Coin Image Acquisition

Image processing and pattern recognition based numismatic studies are mainly focused on coin classification. The very first step in this pipeline is image acquisition. Image acquisition is one of the most challenging and important steps in coin studies as the final results are highly dependent on the quality of the images. Different studies in numismatics need to deal with different challenges. In coin grading studies, the coin image acquisition challenges could be listed as below:

(1) *Resolution*: Coin grading requires images in which the resolution is high enough to include any single default in the coins. Low resolution images usually do not reveal many of the small bumps and scratches on the texture of the coins. On the other hand, very high resolution images will include a significant amount of noise which is not desirable either. Therefore, it is important to investigate and find the required resolution with respect to the study. The dependency of the different numismatic studies on the resolution of the images varies a lot. For example, in coin recognition studies the images can be as low resolution as 100 by 100 pixels, so that the different patterns on the coin images are distinguishable. The acquired images for coin grading and fake coin detection studies should be of high resolution since small details and changes in the texture and the profile of the coins are considered to be important factors in these studies. In coin grading, the issue is more sensitive since even the small scratches and wear effects should be taken into account.

(2) *Luster band effect*: In coin photography the most challenging part is how to deal with the light reflection. Coins minted in different years usually contain different percentages of metals and sometimes even the metal type is changed. Each metal reflects a certain portion of received light, which affect the final image in different ways such as over magnifying some bumps and scratches, underestimation of edges or scratches, and false visualization of background texture as edges or scratches. These effects are much more important when it comes to coin grading, as the base of the study is concentrated on the amount of wear and the severity of the defaults on the coin.

The factors which should be considered carefully in coin photography are listed in [28] and summarized below:

Exposure

The amount of light allowed to touch the camera sensor during the photo shooting. Darker scenes need longer exposure.

Camera Lens

The two main specifications of the camera lens, *focal length* and *aperture*, affect the quality of the images directly. The focal length is referred to as ”the distance from the optical center of the lens to the object” and aperture ”limits the amount of brightness of the images” [28].

International Standards Organization (ISO) Sensitivity

ISO in cameras sets the sensitivity of the camera sensor to the light. Darker scenes need a higher ISO setting; when the high ISO setting is used, more noise will also be captured. Therefore, a wise choice of ISO especially for coin photography is recommended.

Camera Stand

In order to have a uniform dataset and fine circular images of the coins, a stand is required in coin photography. The stand should be set up at a proper distance from the coin, so that the best result in terms of details of the image is acquired.

Illumination and Supporting Surface

Light setup has a significant impact on coin photography. Due to the metallic surface of the coin, reflections and handling them are challenging. Moreover, in one coin (especially the Toonies, which are the targets of this study) different types of metals are used. The amount of light reflection from each metal varies from one to another. Therefore, the lighting system should be designed to minimize the amount of reflection and give the best relative illumination for all parts of the coin. Furthermore, the color of the supporting surface should be chosen so that it neither casts shadows and nor affect the camera's light calibration.

2.2.1 Dataset Specifications

The dataset created for this study consists of Canadian Toonies, which are all gathered and photographed at CENPARMI. Based on the photography details released by numismatic centers as in [29], and trying more than five different setups to create a good set of images for the study, the following setup is proposed for the experiments regarding coin grading.

To achieve the desired image resolution, a ”CANON 60MM F2.8 MACRO EF-S” camera lens is used. Ring light had shown the best outcome when the camera is set to the exposure of 1/25 (s) and ISO of 100. Ring light gives an equal distribution of light on the surface of the coin. Also, the black construction paper that is used as the background will minimize the effect of camera inner calibration that leads to low visual representation of the details. The selected background also eliminates the shadow casting on the outer edges of the coin, and the combination of all these setups will assure that the relative intensity of the different parts are more or less kept, meaning severe magnification or underestimation of the edges/scratches do not exist in the resulting images.

An example of the final resulting images can be found in Figure 2.4a. Comparing this image with a sample from one of the most well-known coin datasets created for coin recognition in Figure 2.4b, one can easily observe significant details which are necessary for this study in CENPARMI photographed coin image. The other difference which is not shown in the images is the size of the image and the coin in the image. The initial resolution of the camera for CENPARMI photographed coins is set to take images of size 2000×2000 pixels approximately, which leads to coins with approximate radius of 700 pixels. This allowed us to have substantial details included in our images, so that the coins could be comparable in terms of texture quality, whereas in MUSCLE or similar datasets the resolution is limited to 640×576 pixels, in which the maximum coin sample radius is roughly about 200 pixels. The final prepared set of images contains 129 images from the obverse

(a) A sample of well-known MUSCLE dataset [30] (b) A sample of CENPARMI photographied Toonies

Figure 2.4: Comparison of CENPARMI and MUSCLE dataset

side of the Toonies. The reverse has almost half of the importance of the obverse in coin grading, and is only included for a more detailed study covering all the aspects of coin grading, which is out of the scope of this thesis. The images consist of Toonies from 2009, 2011, 2013 and 2015. The portraits on these Toonies are slightly different from each other and even within one class, for example 2009, slight changes in the dimension of the queen's head can be found. The intention of choosing similar portrait designs, was to be able to distinguish the small textural differences exist among the different groups of coins by filtering the larger scale changes. However, as for the small changes in the dimension and facial lines of the queens even within these most similar designs the proposed algorithm shall be more biased towards the texture quality.

The prepared dataset is then graded by a coin specialist and the coins are labeled based on their quality. Below an illustration of the different quality levels existing among the dataset, is given.

- **Uncirculated (UC)**: Extremely sharp and shiny edges. The field is smooth all over the coin. The field can be affected by the visible mint-marks. In some cases, several minor scratches might be traced (Figure 2.5a).

- **Choice Extremely Fine (EF+)**: The design edges are a bit smoother than the almost uncirculated coins. The field around the design is almost clear. The field surrounding the legend has some scratches (Figure 2.5b).

- **Very Fine (VF)**: The design edges are severely worn out and although the design is complete, small details are missing from the hair and the neckles of the queen. The field is full of small bumps and scratches and sometimes big and deep scratches can be traced on the coin. Due to the wear some parts of the legend are buried into the field (Figure 2.5c).

(a) An uncirculated sample (b) A choice extremely fine sample (c) A very fine sample

Figure 2.5: Samples of different quality degradation levels exist in CENPARMI dataset

Chapter 3

General Preprocessing

3.1 Problem Definition

Given the explanations in the previous chapters, the problem of coin grading is categorized as a texture detection problem. As mentioned in section 1.4, the coin grading problem is a much broader topic. However, the focus of this study, is on the quality detection and feasibility of having a computer-aided grading system to have a general approach responding to the other applications listed in section 1.2 as well as coin grading. Also, as mention in section 2.2.1, the dataset fall into three main quality categories based on the Sheldon scale (see Appendix A), which for a complete coin grading study, the dataset shall fully cover the different quality categories in sufficient number. Therefore, some elements of coin grading such as ”eye appeal” and ”luster” detection are not covered and the focus is on wear quality detection.

CENPARMI Canadian coins dataset has some challenges to deal with, prior to using them in the main algorithm:

(1) Slight differences in the size and resolution of images

(2) The scale of the coin’s image with respect to the whole image keeps changing from image to image

(3) The edge detection in some images is adversely affected by negative illumination changes

(4) The illumination setups of different images are not exactly the same due to changes in the material of the coins and slight lighting setup changes in each set of data acquisition

3.2 Preprocessing Steps

To deal with the above listed problems, the following preprocessing steps have been considered for this dataset:

(1) Image normalization,

(2) Background elimination, and

(3) Illumination removal.

3.2.1 Image Normalization

Image normalization is the key step in order to make a stable dataset. As the images in the CENPARMI Toonies dataset are nearly similar in size, normalization will not affect the effective features in the images. The applied image normalization process is summarized in Figure 3.1. The following procedure could have been done using the Hough transform to detect the coin and crop the image to the bounding box of the detected coin. However, comparing the performance of both methods, the Hough transform was two times slower than our proposed work flow. As a result, the Hough transform is suggested only for the parts where a detailed answer is required.

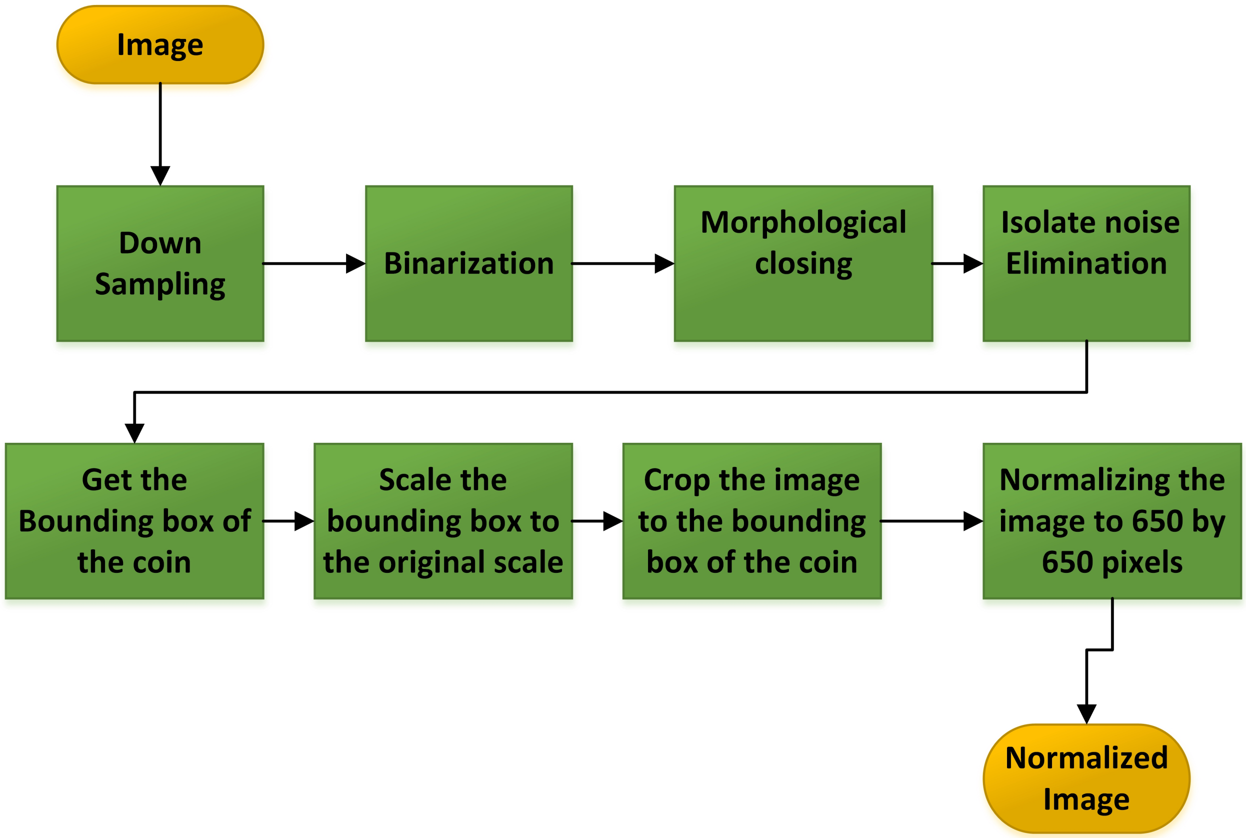


Figure 3.1: Image Normalization work flow

Down Sampling

The original image matrices have about 2000 rows and 2000 columns, which is noticeably large for image processing algorithms. In order to speed up the preprocessing steps, the images are down sampled by a factor of 4, which breaks the resolution down to 500×500. The color images are converted to the gray-scale which contains sufficient information for the following processing steps. The preprocessing steps are implemented on the down sampled gray-scale images, and the final result will be scaled back to the original image size. Figure 3.2 illustrates the down sampled gray-scale image.

Figure 3.2: Comparison between the original and the gray-scale down sampled image.

Binarization

The only goal of binarization in this part is to have a rough approximation of the size and the shape of the coin, as well as facilitate the image morphological processing to be done in the following steps. As in this step, we do not care about the inner edges of the coin, any edge-image that extracts enough information concerning the outer boundary of the coin while keeping the performance of the preprocessing is a good choice. Therefore, Canny edge detection is used as it is one of the fastest algorithms to get a binarized image. The level of details is adjustable by changing the parameters of the Canny to give the freedom to have as much detail as possible. This is important since due to the negative effect of illumination in the images, other edge detectors simply neglect some parts of the boundary. On the other hand, Canny leaves it up to the user to decide how much detail should be kept, and this will lead to having an almost connected outer boundary when a large amount of detail is preserved in the images. This boundary then can easily get completely connected using an appropriate structural element and morphological closing operation. The result of Canny is illustrated in Figure 3.3a.

(a) Binarized image using Canny

(b) Closing the gaps in the outer contour

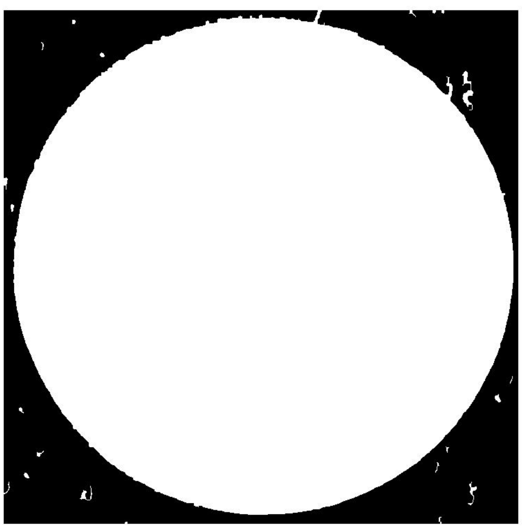
Figure 3.3: Extraction of the outside boundary of the coin

Morphological Closing

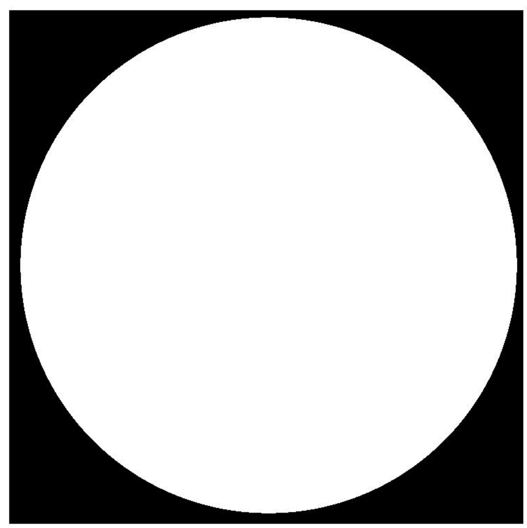
The outer boundary of the coin is not fully connected, and any attempt to use shape detection will end up only filling the holes which are fully connected. Using morphological closing with a disk shape structural element of size 3 pixels, the outer edge will be connected. As mentioned earlier, we do not care about the effect of closing on the inner edges at this point, as the important factor is to have a fully connected outer contour to extract the bounding box of the coin.

Elimination of Isolated Noise and Extraction of the Corresponding Mask

Using the outer boundary of the coin, the inner region of the coin is completely filled by 1 in the binary image. This will create a mask filtering everything other than the coin. However, some isolated noise can be found in Figure 3.4a, which should be removed to have a perfect circular shape mask. With the help of connected components detection and knowing that the isolated noises are the connected components with less than 500 pixels, the connected components are detected and this noise is eliminated. Finally, a clean mask can be extracted as in Figure 3.4b. To remove the zigzag pattern from the outer boundary edge of the mask, erosion is applied. A disk shape structural element of size 4 pixels is used in this step.



(a) Morphological Filling



(b) Noise Removal

Figure 3.4: Coin mask extraction

Crop the Image to the Outer Boundary of the Coin

Now that the clean mask is extracted, the bounding box of the shape is detected (Figure 3.5a) using the minimum and maximum coordinates of the circle in each direction. The bounding box size is then scaled by the factor of 4 to be fit to the original image and the gray-scale image will be cropped with respect to the scaled bounding box (Figure 3.5). The last step is to normalize the gray-scale image to 650 $\times$ 650 pixels. This makes sure that the size of the coins is kept the same in all the images. Therefore, it is easy to set the Hough transform configuration only for one coin and apply it to the whole dataset for the next step of preprocessing. The whole procedure is done so that it can be transmitted to the colored image, and have the colored image as the output in case of the need for further investigations on color images.

(a) Extracting the bounding box of the mask (b) Scaled bounding box to the original 2000×2000 image (c) cropping and scaling down to 650×650

Figure 3.5: Final result of normalization

3.2.2 Background Elimination

Background elimination means to keep the target pattern in the study which is also called foreground and eliminate all other patterns which are also called background. This step is required in order for the texture study to be concentrated on the coin and its variations. Although in the photography step a dark background was chosen, it still contains slight variations of light which could be problematic in edge detection and feature extraction steps. In this step, the coin extraction shall be more precise as the details of the foreground shall be kept. Below, the steps chosen to do the task are demonstrated:

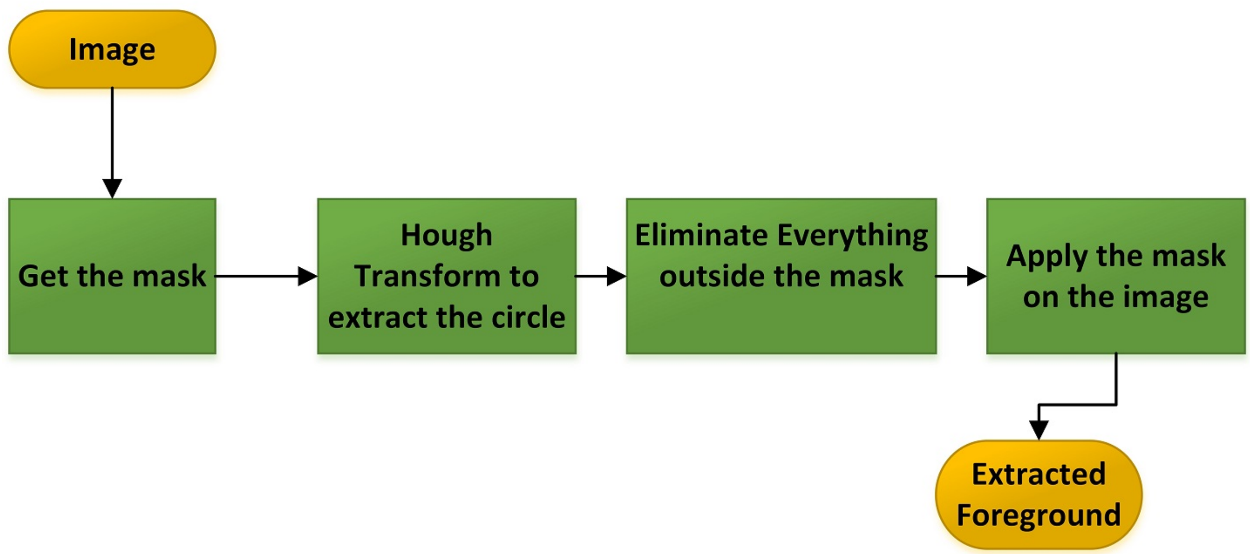
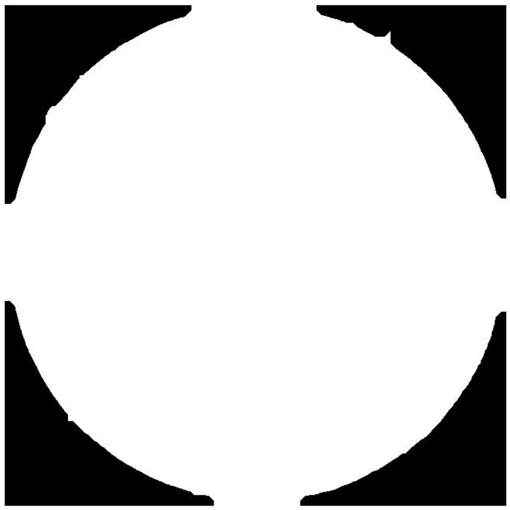
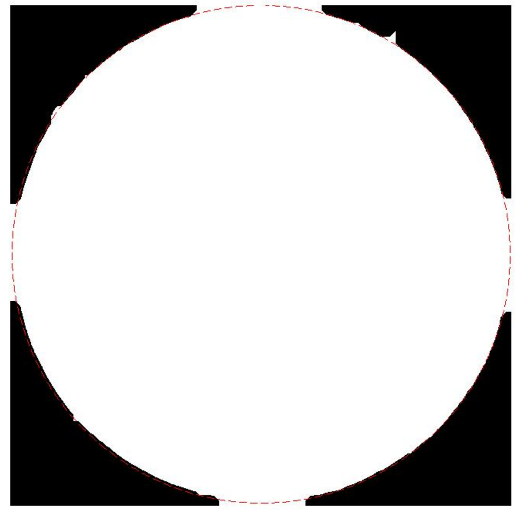


Figure 3.6: Background elimination process

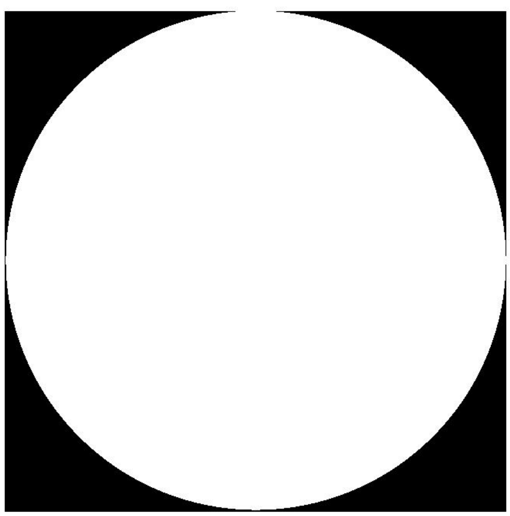
The same mask extracted in the previous section is used for the first step. The reasoning behind having another Hough transform step is that sometimes the mask is not a fine circle as the illumination effect on the outer contour of the coin is too high and the morphological approaches can be misleading. The Hough transform is a mathematical algorithm which is not affected by the photographic parameters. In this algorithm, which is based on a three dimensional voting system, the centroid and the radius of the circle are estimated. Based on these estimations, the circle is easily extracted. Normalizing the images prior to background removal will help to have a closer estimation for the coin radius in the images. As per previous discussion, all the images are cropped and normalized to 650×650 so that the image's edges touch the coin's outer contour. Therefore, the radius of the coin is approximately 325 pixels. This rough estimation reduces the computation cost of the Hough transform by decreasing the number of tries in the algorithm to find the radius with maximum number of votes. In order to have a precise approximation, it is usually recommended to use the edge-image and apply the voting system on the binary image. We also save the computation cost, by using the circular mask instead of the edge-image. The circular mask is a rough estimation of the coin and since that is the only circle we are looking for, keeping so many misleading edges is careless. The initial image and the final image are shown in Figure 3.8 and the step by step practice can be followed in Figure 3.7.



(a) Detected mask with morphological processing



(b) Detected circle using Hough transform



(c) The fine circle mask

(d) Foreground pattern

Figure 3.7: Background removal

When the coin's circle is detected everything outside the detected circle is set to zero and a precise circle shape mask is created. In this step, one can easily eliminate the background by applying the mask on the image. Finally, the image is cleaned up, meaning all the illumination effects on the background and outside the coin are removed to produce an image which only contains the coin.

(a) Original normalized image (b) Background removed image

Figure 3.8: Background removal

3.2.3 Illumination Removal

As discussed earlier in this chapter, changes in illumination in different areas of the image are not desirable. Such changes may cause several undesirable effects on rating the coin based on the images.

(1) *Underestimation of some edges* when the illumination on the ”field” of the coin is changing. This may cause too much resemblance of the ”field” to the ”legends” or ”portrait” and mislead the edge detector incorrectly identifying the main edges. It may also cause the scratches to not show up as strong in some cases. An example of this effect is illustrated in Figure 3.10a. The legends in the left side of the coin have stronger edges whereas they are not well defined in the right side.

(2) *Over-magnification of some edges* is another issue as a result of reflection. An example of this effect is demonstrated in Figure 3.10b in which the field texture is represented similarly to heavy scratches. The coin itself is a mint-state coin on which the die marks are noticeable,

however, due to the misleading illumination effect, the marks are represented like heavy scratches. It is also noticeable that the legends are not bold, although the coin is a mint-state.

(3) *The coins' base metals and the portion of used metals keep changing every year*. Sometimes the change is noticeable in the colors of the metals. This effect in coin photography and in gray-scale images is represented by a different intensity level. Having a larger difference between the intensity levels of the legends and the surrounding field leads to having a cleaner edge image. This causes the coin in Figure 3.10c to be graded as a higher quality level than the one in Figure 3.10b, although they are both categorized as uncirculated coins.

Due to the listed effects, it is suggested to remove the illumination when the algorithm is not illumination invariant. The designed approach to eliminating the illumination changes from the images is depicted in Figure 3.9.

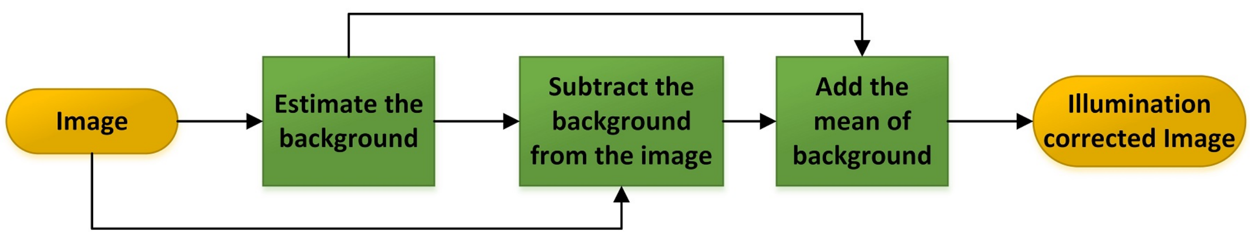


Figure 3.9: Illumination Elimination process

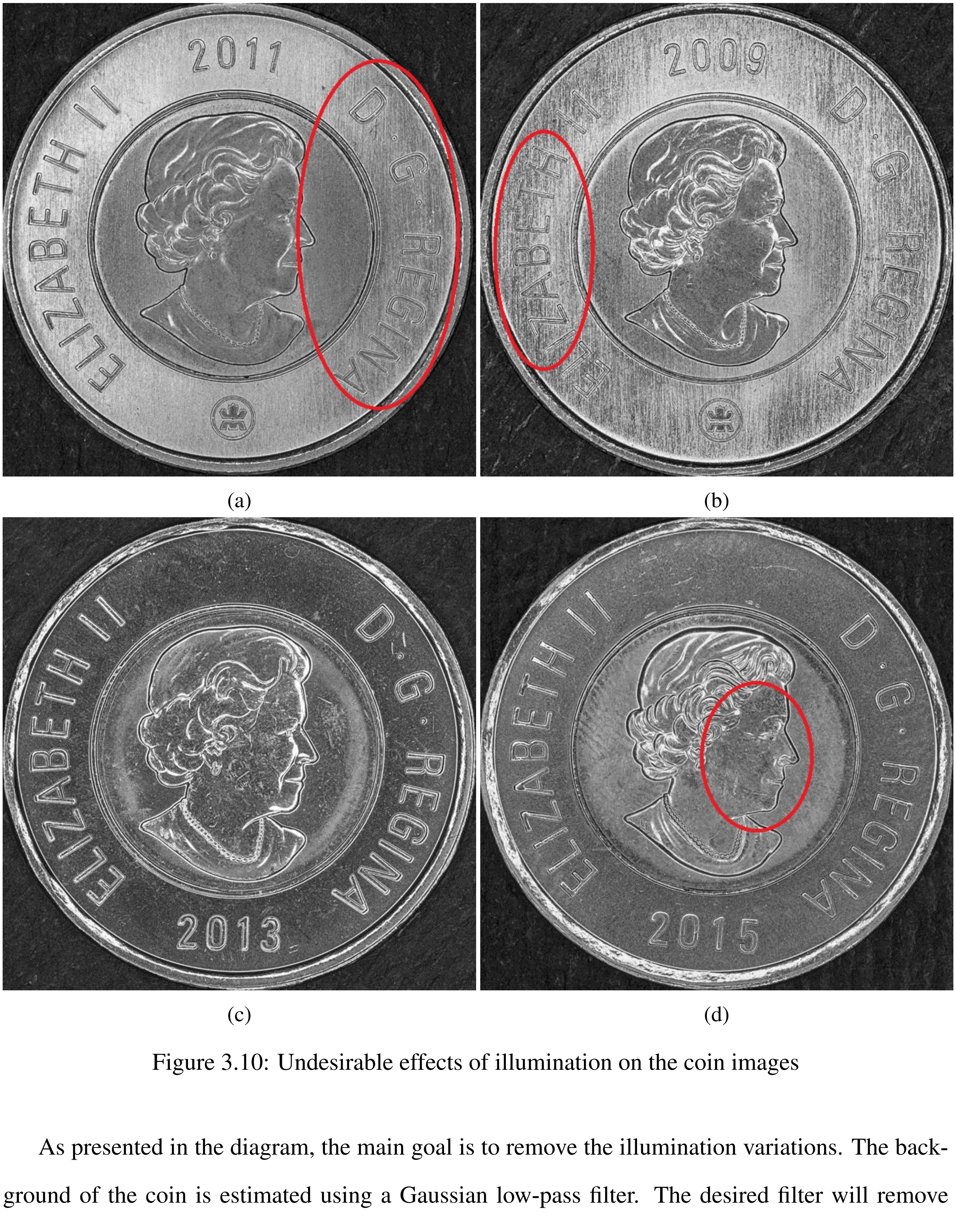


Figure 3.10: Undesirable effects of illumination on the coin images

As presented in the diagram, the main goal is to remove the illumination variations. The background of the coin is estimated using a Gaussian low-pass filter. The desired filter will remove all the details from the image and will only display a vague image which contains the intensity variations in different areas. The sigma value for the used Gaussian filter is set to 15 and the Gaussian function is sampled in a window filter of size 61×61. Removing this image from the main image will reveal the high points that we would like to keep, since the information of the image is saved in these points. In order to have a smooth background, the mean value of the estimated background image is used to be added to the high value image as the new background. As a result, the illumination changes in the background will be removed to a good extent and at the same time the important information of the image will be kept. The result of such procedure is represented in Figure 3.11.

Figure 3.11: Result of illumination correction

Chapter 4

Traditional Image Analysis for Coin Grading

4.1 Introduction

In the past decade, machine learning algorithms have emerged and evolved fast. They have found their place in almost all of the existing research studies and have become the main assessment of almost all studies in just a few years. This is even more tangible in image processing and pattern recognition studies as the learning algorithm is an inseparable part nowadays. However, we shall not forget that for years image processing and pattern recognition studies were done without the help of learning algorithms. Pattern recognition and image processing, themselves are powerful tools to analyze the images and for every study to be legitimate, there is a need to investigate the traditional approaches. In this section, we propose an image analysis approach for a rough estimation of the coin's surface wear. As discussed in chapter 3, the images are already corrected with respect to scale and translation. However, the rotation changes was not in the scope of the previous sections to be fixed as the proposed method was rotation invariant. In this chapter, first the angular differences within the images are fixed to achieve rotation invariance. Using the edge image and the registration techniques the scratches, bumps and defaults on the coin's surface are estimated.

4.2 Wear Detection Algorithm

The flowchart depicted in Figure 4.1 shows the steps for this approach. The main difference between registration techniques and feature extraction techniques (which an example of that discussed in chapter 5), is that the small changes which were negligible in the previous section such as the year's spot on different coins, and different obverse patterns are of high importance in registration and should be prohibited. For this reason, the steps below are repeated for each group of coins from 2009, 2011, 2013, and 2015 separately.

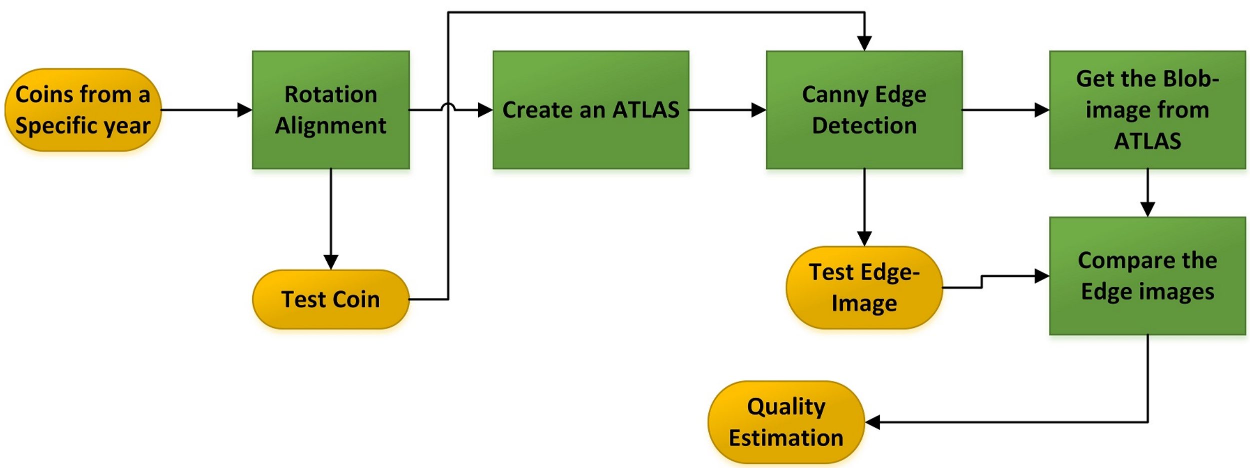


Figure 4.1: Wear detection using image registration

4.2.1 Rotation and Alignment

Correlation and Registration

There exist so many different methods to align two coins. The common approach is to use correlation to detect the rotation angle. First a template coin is selected. A test coin, is rotated each time by a fixed degree and the correlation between the test image and the template image is calculated. The rotated version of the test image which results in the highest correlation, is selected and the rotation angle is recorded. The coin images captured in CEPARMI do not have more than 3 degrees of angular difference due to the precise setup used in the image acquisition step. Therefore, for each group of images belonging to the coins from a specific year in the dataset a sample coin is selected and the remaining coins from that year are compared to the sample coin by only rotating them up to 3 degrees in both clockwise and counter clockwise directions taking 0.1 degree steps. Since the coin's design has a lot of details, it is suggested that the correlation is calculated only on the legends. Therefore, the inner circle of the coin is discarded from the correlation computations. Also, the edge image gives a better response in this step since the illumination is not constant for the entire set of coins, the calculations on the gray-scale image leads to an increase of the error rate. Due to the small pattern variations even in the coins from the same year, this approach failed to give the best rotation in some cases. Therefore, a more stable method is implemented.

Matching Key Points for Alignment

Key points are strong stable points of the image which can be used for extracting objects in different images and match them together. SIFT algorithm (see section 5.2.1) is a powerful method to extract stable key points and chances that these key points fail to match the right key point in the other image, is very low in CENPARMI dataset. However, it is not rational to do all the process of detection and matching the SIFT key points for aligning the coins. Speed-Up Robust Features (SURF) are the speed up version of SIFT. These features use Hessian matrix for localization which is faster than SIFT, however it has the disadvantage of being rotation invariant for only up to 15 degrees angular difference between the images. Since the maximum angular difference in our dataset is 3 degrees, this is not a problem. Again a template image is chosen for each group of coins. The SURF key points and their locations for each of the template and the test images are calculated. The key points are matched between the test and template images using the Euclidean distance. When all the key points are matched, the rotation matrix that results in the new locations of the key points in the test image, is calculated. The inverse of the calculated rotation matrix is then applied on the test image to rotate and align it with the template image. Since this method is focused on a subset of pixels in the images which have the most important information, it is not misled by the illumination changes and small differences in the patterns, and therefore the resulting images have been satisfying.

4.2.2 Creating an ATLAS Image

In registration techniques, it is very important to choose the right template to which all the other images are compared. As discussed in chapter 3, section 3.2.3, even images of the mint-state coins are not considered as good templates because the mint struck marks are usually magnified and are represented similar to the scratches. This leads us to create an ATLAS image as the template (master) image. In medical image processing, it is usually hard to acquire images from certain angles or images with certain illumination pattern. In such cases, an ATLAS image is created which is the result of combining different images from the same object so that the resulting image has the required specifications. In this study the required specifications of an ATLAS image is to have an almost fine texture together with clear edges on the design and the legends. In order to create such an image:

(1) All the images should be aligned (reasoning behind the previous step)

(2) Averaging on the aligned images will result in a fine template image

The resulting ATLAS image has:

- sharp and clear edges on the main design, the legends and the rim
- almost a fine and smooth field with no scratches and smoothed struck marks
- corrected illumination changes to an acceptable extent

Figure 4.2 shows an uncirculated image from 2009 and the created ATLAS image for this year. As it can be seen, the field marks on the uncirculated coin can be estimated as possible scratches making it not a good choice to be the master coin. However, the ATLAS image has a very fine field texture and sharp legend edges which are desirable to be selected as the master coin.

(a) Uncirculated coin image (b) ATLAS image

Figure 4.2: Comparison between an uncirculated coin and a created ATLAS image

4.2.3 Create a Blob Image from the ATLAS

The whole idea is to compare the ATLAS image and a test image to look for the texture defaults and estimate its wear quality based on this comparison. In this procedure good edge images are required for both the ATLAS and the test image. An edge image is considered to be good if all the main edges are detected and there is no false positive answers (i.e. a detected edge which is not an edge in the source image) resulting from intensity changes. That leads to an almost clean edge image for the ATLAS and an edge image showing all the bumps and scratches on top of the main edges in the test image. Among the existing edge detection methods, Canny is the best choice for this study since the different parameters can determine the amount of details required. The higher threshold of Canny is set to 0.2 and the lower threshold is the scaled of higher threshold by the factor of 0.4. In order to deal with the small changes in the profile patterns, the ATLAS edge image is not directly used in the next step. Instead, the blob image is extracted using morphological closing and a disk structural element of size 3 to close all the open edges. This is followed by a dilation process with a disk structure of size 6 to have a blob image. In the last step an opening process is implemented on the image to remove the unnecessary connections. Figure 4.3 indicates the clean edge image calculated using the ATLAS image and the created blob image.

(a) Result of Canny on the ATLAS image

(b) Coin with smooth surface

Figure 4.3: Creating the blob master image

4.3 Wear Estimation

To estimate the wear and quality of the coin, all the details including bumps and scratches are important. The edge image of a test image is calculated using Canny with the same parameters as in the previous step. The edge image is then compared to the created blob image and the defaults are extracted. An estimation of the wear is then given to the coin based on the number of regions and the density of each region. The bumps are usually small connected components which are distributed over a large area on a heavily worn out coin. The scratches on the other hand are concentrated on a small region but they have a dense group of pixels creating them. Figure 4.4 and 4.5 show the result of a heavily worn out coin and an almost uncirculated coin after comparing with the blob image extracted in the previous part and the defaults of each coin are extracted and illustrated.

As demonstrated in Figure 4.4, the blemishes in a heavily worn out coin are distributed all over

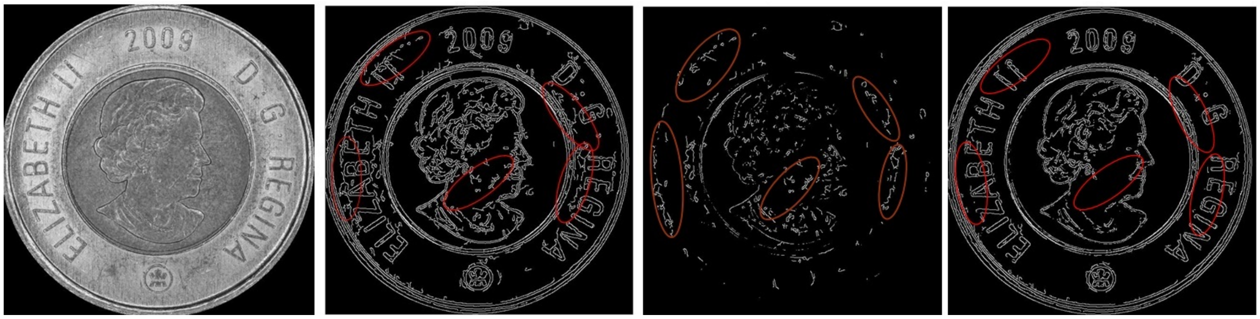


Figure 4.4: Extracting blemishes from a worn out coin

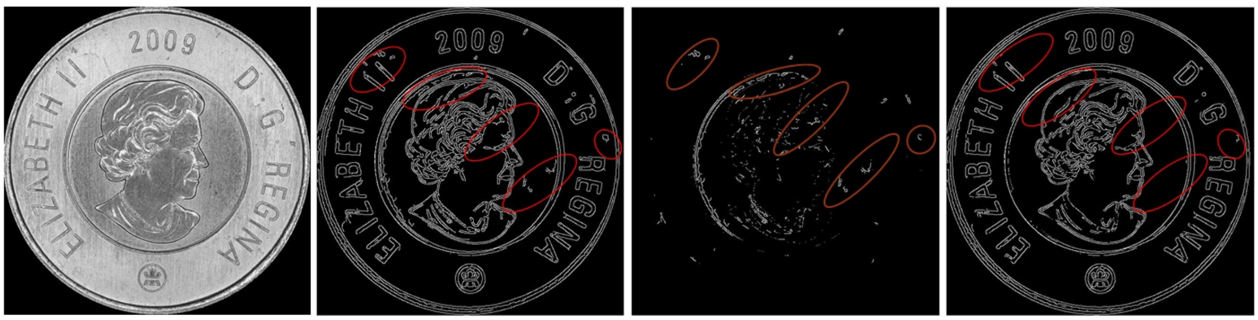


Figure 4.5: Extracting blemishes from an almost uncirculated coin

the coin's surface which is an estimation of the wear on the field. It also contains certain regions of densely concentrated pixels which are the scratches. On the other hand, Figure 4.5 shows that an uncirculated coin has almost a fine field. It is categorized as almost uncirculted rather than uncirculated only because of certain dense regions which show some scratches on the surface. Therefore, this method can differentiate between an uncirculated, almost uncirculated, an a worn out coin by assigning a scale to the coin based on its wear density. Both Figures start from the original image (most left), edge-image, the extracted belimeshes from each image, and the final image shows the result of reduction of the blemishes from the edge-image. As it is expected, the final cleaned edge-image of a worn out coin is very different from its original edge-image, whereas for the case of almost uncirculated coin, the differences are not noticeable.

In order to have a quantitative measure for what just discussed, one shall notice that the blemishes are getting more important as they get further from the design. This is because this method is an edge-based estimation of the wear and in some cases the blemishes on the design might not be the actual blemishes on the coin. Therefore, as we get further from the design the chance of having the actual bumps and scratches extracted is higher. Considering the two coins in Figure 4.6, the procedure is explained on the images of an uncirculated (left) and a very fine coin (right).

Figures 4.6a and 4.6b clearly show the differences between the quality of an uncirculated and a very fine coin. The edge extraction results using canny, are shown in Figures 4.6c and 4.6d respectively. After comparing with the blob image, the extracted blemishes from each image are represented in Figures 4.6e and 4.6f.

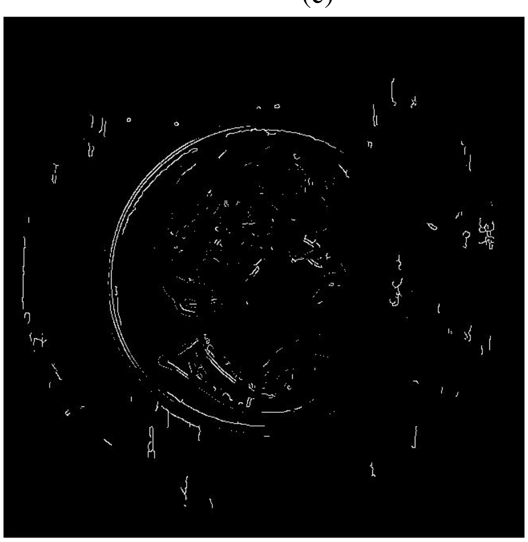
The images illustrated in the last row of Figure 4.6, contain some noise as well. Therefore, in the next step the connected components containing less than 1000 pixels are rejected from both of them. The remained components are then weighted by the density of the pixels in each component. However, the wear needs to be inversely weighted with respect to the distance from the center of the image for the sake of making sure that the field wear is given a higher importance than the wear in design area. We used a Gaussian weighting kernel in this step. Having the weighted pixels now, the sum of all these pixels are calculated and the result is normalized in the scale of 0 to 1, to give a quantitative grade for the amount of detected wear on the coin. Using this method, the uncirculated coin receives a grade of 0 (meaning no wear is detected) and the very fine coin receives a grade of 0.8386.

(a) Uncirculated coin

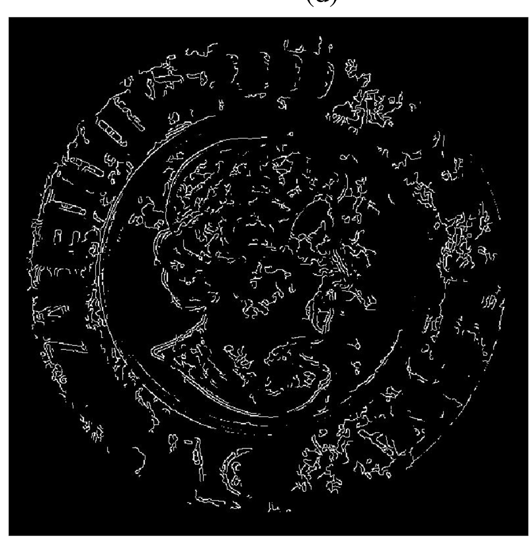
(b) Very fine coin

(c)

(d)



(e)



(f)

Figure 4.6: Extracting the wear on an uncirculated (left) and a circulated (right) coin

4.4 Limitations

This algorithm provides a good understanding of the overall amount of blemishes on the coin with respect to the other existing coins. However, the limitation of this method led us to choose a more reliable algorithm provided in the following chapters. The limitations of this algorithm are listed as:

(1) The resulting wear grade will change if more heavily worn out coins or coins with different levels of quality are added to the existing collection.

(2) The algorithm is not design invariant and therefore, the procedure must be repeated for each set of coins with a specific design separately.

(3) This algorithm does not provide detailed grading of coins with high quality (i.e. the differences among the uncirculated and almost uncirculated coins are hard to detect with this algorithm). Since, the algorithm is not biased with respect to the wear on the main edges of the coin.

Chapter 5

Application of Texture Analysis in Coin Grading

5.1 Preliminary Data Analysis

5.1.1 Dataset Labeling

As discussed in chapter 1, the process of grading a coin is very subjective. In human grading, usually three graders' judgments on a coin are used before stating the final label. It is also worth mentioning that not all the parameters mentioned in previous sections have the same weight for grading the coins. The goal of this study is to create an algorithm which has a similar judgment ability of a coin specialist, hence it is important to understand the coin specialist's mindset. In order to fulfill this prerequisite, with the help of Dr. Suen, we were able to find a coin specialist who carefully investigated and labeled our dataset. The main factor that coin specialists consider in grading the coins is the overall amount of wear, scratches, bumps and also the wear of certain defined high points for each type of coin. Based on the listed parameters, the problem is clearly categorized in the texture analysis type of study. In the following subsection, the preliminary texture analysis attempts are described. The results from this section, helped us to understand the small differences among the data classes and lead us to choose a compatible set of features for this study which is described in details later on in section 5.2.

5.1.2 Principal Component Analysis Algorithm

Principal Component Analysis (PCA) is a statistical technique widely used in both image recognition/classification and data compression. The main goal of PCA is to describe the data in terms of certain unit vectors which magnify the variance within the data; in other words, PCA detects the patterns in the data so that their differences are highlighted. Along with this purpose, it also provides dimensionality reduction [31]. In order to find a representation of data which signifies the variance, PCA uses the two concepts of covariance and eigenvectors.

Covariance

The covariance between every two features is defined as below:

$$cov(x,y) = \sum_{i=1}^{n} \frac{(X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)} \tag{1}$$

In the above formula, X_i and Y_i are the i$^{\text{th}}$ values of features x and y. $\bar{X}$ and $\bar{Y}$ are the data mean values. As it is well-defined, the standard deviation and variance are measurements of how spread out the data is in a one dimensional space. When it comes to more than one dimension (i.e. more than one feature), covariance is the key to measuring the correspondence between every two features. The most important thing about the covariance value is it's sign. The negative sign indicates that the features are inversely related, whereas the positive value means they are directly related. A zero value for covariance means that the features are independent from each other.

Eigenvectors

An eigenvector is defined for a transformation so that when the transformation is applied on eigenvector the resulting vector is a scale of eigenvector. For an $n \times n$ size matrix, n eigenvectors could be found. All the eigenvectors are perpendicular to each other, which could create the basis vectors of another space, meaning the data could be expressed in terms of these vector. The value by which the eigenvector is scaled when multiplied to the transformation is called the *eigenvalue*.

In order to find the principal components which make the most significant differences within the dataset, the following steps should be taken:

(1) Subtract the mean

(2) Calculate the covariance matrix

(3) Calculate the eigenvectors and eigenvalues of the covariance matrix

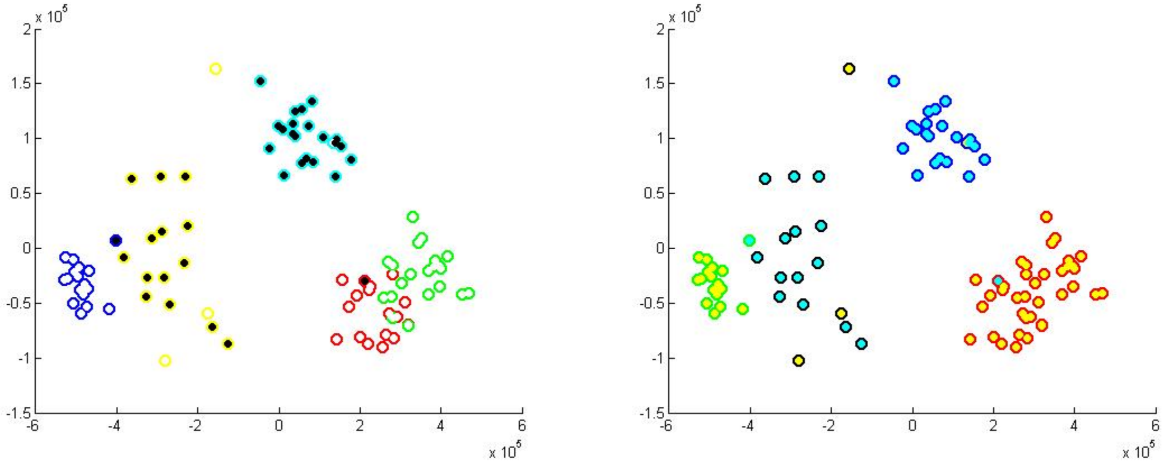
(4) Choose the components and format the dataset

(5) Extract the feature vector

5.1.3 Coin Dataset Analysis and Classification

As discussed in the previous section, PCA is a powerful data analyser which can determine the dominant components of the data that produce the maximum variance within the dataset in a new space. In this step, the dataset is studied and using PCA the dominant components are recognized. These dominant components are then used to classify the data into two groups of uncirculated (UC) and circulated coins. As described in section 2.2.1, the dataset is already labeled in three different quality classes. The characteristics of the classes are fully described in the same section. In this preliminary test, the goal is to first verify whether the differences between the texture of the uncirculated coins and the two other classes, referred to as circulated coins in this section, are bold enough to be recognized without using a proper feature vector. The sum of gray-scale intensities for each row and each column of the image is calculated and the vector which contains these values is fed to PCA for each image. Based on the result of principal component analysis, 86% of the variance among the images fall in to the first two components; however, to make sure having an accurate clustering result and not losing too much information, 40 out of 96 components were used to group the coins. These components were then used to classify the data using support vector machine with a linear kernel. The accuracy in this step was 60%. Noting that the PCA is not invariant to rotation, scale, translation and illumination, the next step to improve the result is to correct the illumination changes in the dataset. As discussed in chapter 3, the images are corrected with respect to scale and translation. Also, due to the camera setup and data acquisition process, the images are almost aligned within the range of maximum 5 degrees error. This small rotation difference is corrected using the registration technique explained late in chapter 4.2.1. The dataset is then corrected with respect to illumination changes as discussed in section 3.2.3. Testing the above algorithm on the illumination and rotation corrected images, improved the accuracy to 81.8%. Therefore, a rough texture analysis can easily differentiate between an uncirculated and a circulated coin.

In the next step, the 40 dominant components are used as the input of an unsupervised learning algorithm as k-means. The goal of this test was to understand whether the computerized methods could be efficient enough to respond to one of the most subjective areas of numismatics which is coin grading. To visualize the clusters, only the first two components are used, due to facilitation of the visualization and knowing that these two components are the dominant elements to define the distribution of the dataset in the new domain. Figure 5.1 shows the result of grouping the 96 training sets in to different numbers of clusters. Figure 5.1 illustrates the dataset being classified in different number of clusters and described with respect to its first two principal components. The cyan and yellow colors in Figure 5.1b show the coins which were classified as circulated and uncirculated respectively. The same definitions have been shown in black and white in Figure 5.1a. The different outer edge colors of circles show the different clusters recognized by k-means. In this concept, the misclassification is defined as a circulated coin classified with the uncirculated coins, and vice versa.



(a) clustering into five groups

(b) clustering into four groups

Figure 5.1: Results of clustering the coins in groups using PCA and k-means

Based on the size of the training set, an attempt is made to categorize the coins into 3, 4 and 5 clusters. After testing each clustering result for 10 iterations, the average number of misclassification for 5 clusters was 15 and the average number of misclassification for 4 clusters was 4 and 5 misclassification for 3 clusters. This was an interesting outcome to compare with the result of coin specialist's labeling. The coin specialist could group the data into 3 main groups identified as "UC" (Uncirculated), "EF+" (Choice Extra Fine), and "VF" (Very Fine). This would not of course change the fact that the whole dataset could be categorized in 3 main groups based on the quality and the texture. A low misclassification error for 4 clusters, is related to the problem of illumination in the dataset. As explained in section 2.2, small variations in the setup and material of some coins made their intensity range to be very different than others. A result of 4 class is related to that set of uncirculated coins which had different intensity range, since the feature vector selected in this part is not stable with respect to illumination changes, the 4 number of clusters is considered as a measurement error. As a conclusion, computer-aided programs, if designed properly, are fully capable of recognizing the quality and categorizing the coins in a more accurate and reliable manner which is not dependent on the human misperceptions and errors.

5.2 SIFT Key Points as Wear Detectors

As fully described previously, a proper texture descriptor feature set is the main requirement for quality evaluation of the coins. The primarily needs for the selected features are to be rotation, illumination and scale invariant.

Our initial dataset was created under a fixed setup in which the size variation among the images was limited to 400 pixels. Hence, the preprocessing step of scale and translation correction to have images with exact same size of 600×600 pixels did not negatively affect our data stability. In cases, where the data is not created with such a precision, the scale correction might have the effect of losing the resolution and precision in some images when image stretching and interpolation is required.

Due to the same reason, rotation correction is not recommended in coin grading studies. Rotation correction involves finding the rotation angle and creating a rotated version of image with interpolation techniques, which again in coin grading studies even small changes in the textural details will affect the result of the grading. As mentioned in the beginning chapters, our data is acquired under a fixed direction and the maximum angular difference among the images is 3 degrees.

As our data is acquired specifically based on the requirements of this thesis, having a rotation, scale invariant feature set is not an asset. However, this study is the base of a new challenge in numismatic studies and hence, the generalization is the key to have a valid and stable algorithm. Therefore, the mentioned requirements, drive the attention to the Scale Invariant Feature Transform technique (SIFT).

5.2.1 SIFT Algorithm

The Scale-Invariant Feature Transform (SIFT) introduced by Lowe [32] has been successfully used in a wide range of image processing studies. These shape descriptors have been used in many studies related to face recognition [33][34], landmark and objection detection [35][36]. The scale and rotation invariant characteristics of these features have been the main reason for their popularity since 2004, and in many cases they have boosted the result of classification noticeably. In recent studies, applications of SIFT descriptors on texture analysis problems have been investigated and showed promising results [37][38][39]. We propose using the SIFT descriptors for coin quality estimation and grading purposes. The nature of this study, is different from other textural studies as the feature points are variable even among the images within the same class. In the rest of this chapter, we focus on the explanation of the SIFT implementation and how it is a proper feature for our purpose. The following diagram shows how the SIFT descriptors are extracted from a given image.

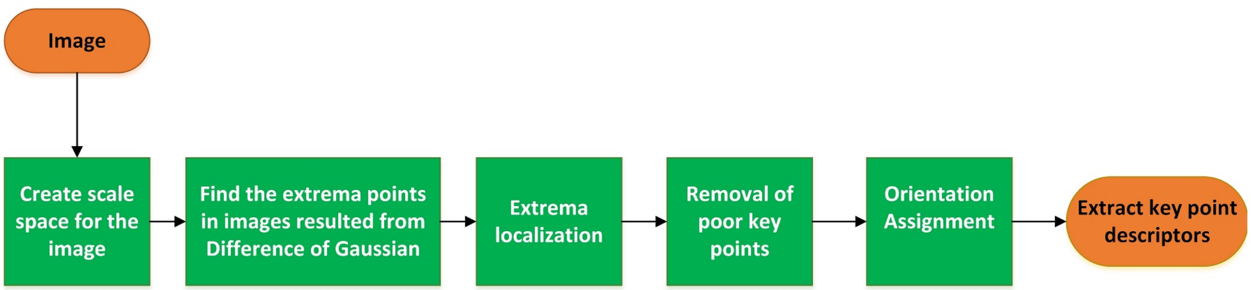


Figure 5.2: SIFT Algorithm Diagram

Based on the depicted diagram, the algorithm consists of several steps:

(1) **Creating a scale space**

The goal of having a scale space is to simulate the effect of changing camera focus while taking pictures from the same scene and therefore providing the requirements of extracting scale-invariant features. The creation of scale space is described as progressively blurring out the image using a Gaussian filter and by progressively, it means that at each new scale the σ parameter in the Gaussian is multiplied by a constant value (similar to the effect of zooming out the camera). SIFT algorithm takes the scale space construction to another level by continuously resizing the original image to half of its size and creating the scale space for each new image. The common practice is to start from the double sized image. Each of these resized images, is called a new octave and at each octave the image is progressively blurring out to create the different scales of that image. Lowe [32] suggests that the best outcome of the SIFT is given using 4 octaves and 5 scales, which yields to a total of 20 images. However, the number of octaves and scales is dependent on the size of the original image.

(2) **Finding extrema points**

The octaves are formed from blurred out images using Gaussian filters with different scales. In order to detect stable extrema points, and save the computation cost at the same time, difference of Gaussian images are calculated using the difference between two consecutive scale images [32]. The Gaussian kernels in the previous step, work as a blob detectors, and with changing the σ value, they try to find the correct scale in which the extrema points could be extracted. Equation 2 is the formulation of Gaussian kernel in which σ is the scale parameter. The difference of Gaussian images are calculated as in Equation 3. Here, the scale parameter of the next blurred image is k times of the previous one. G is the Gaussian kernel described in Equation 2 and I is the original image. D is the calculated Difference Of Gaussian (DOG). Note that, having 5 scales in each octave, we can generate a total of 16 (4 $\times$ 4) DOG images which is enough to locate the extrema points.

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{(-\frac{x^2+y^2}{2\sigma^2})} \tag{2}$$

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \tag{3}$$

When the DoG images are created an extrema point is detected comparing every pixel with its eight neighbors in the same image (space invariant) and with its nine neighbors in the previous and next scale (scale invariant). We do not look for extrema in the lower most and topper most scale images.

(3) **Extrema localization**

Once the key points are detected, their locations are estimated in order to have more accurate results. Using the 3-dimensional Taylor expansion described in Equation 4, the closest estimation for the location of the key points is calculated. This is also called subpixel estimation.

$$D(X_0 + h) = D + \frac{\partial D}{\partial X}^T \Big|_{X=X_0} h + \frac{1}{2} h^T H(X) h \tag{4}$$

In above formula, D is the DOG value in the location of $X_0 = (x_0, y_0, \sigma_0)$ and h is the required offset to find the subpixel location.

(4) **Elimination of poor key points**

Having a stable set of key points in a strong pattern recognition algorithm is an asset. In order to make sure that the detected key points are stable, the low contrast key points and the edge points are eliminated from the set. The edge points are not considered stable, since even if not located properly, Difference of Gaussian has a very big response in these locations. The reminder is the set of key points with strong interest. In order to do so, the gradients of the key points along x and y directions are calculated. If the area around the key point has no variation, the values of both gradients are small. This is the case of having low contrast key point. If the key point is located on an edge, the value of one of the derivatives is large and the other one is small. At last, the interesting key points are located, where the gradient response in both directions is large.

(5) **Orientation assignment**

An orientation should be assigned to each key point to satisfy the rotation invariant claim of the algorithm. To do so, a histogram of gradient magnitude in a window around the key point is created. The size of the window is defined considering the scale, the bigger the scale, the larger neighborhood around the key point shall be considered. The histogram contains 36 bins to cover the 360 degrees and is weighted by the magnitude of gradient with respect to each direction. The peak of the histogram is set as the key point's orientation and all other bins with 80 percent of the peak are considered as the new key points. Therefore, in the same location and scale more than one key point with different orientations could be recognized, which makes the algorithm more stable. The formula used to calculate the magnitude and the orientation of the gradient for each neighbor pixel are shown in Equations 5 and 6. The notation L is used for the scale images calculated in the first step by convolve with the Gaussian kernels.

$$m(x,y) = \sqrt{[L(x+1,y) - L(x-1,y)]^2 + [L(x,y+1) - L((x-1),y)]^2} \tag{5}$$

$$\theta(x,y) = \arctan\left(L(x,y+1) - L(x,y-1)\right)/\left(L(x+1,y) - L(x-1,y)\right) \tag{6}$$

(6) **Extract key point descriptors**

Finally, to describe the key points in terms of scale, rotation and illumination invariant features, a 16×16 neighborhood of the key point is selected. The magnitude and the orientation of gradient for each pixel in this block is calculated. To achieve the rotation invariant, the key point's orientation is subtracted from all calculated gradient orientations in the big block. It is then divided into subblocks of the size 4×4 and an 8-bin orientation histogram is calculated for each subblock. The histogram calculation is similar to the previous step, however a Gaussian kernel is multiplied to the main block prior to the creation of histograms, to reduce the effect of the neighbors which are far from the key point. Adding the calculated histograms for each sub-block, one after another in a vector of size 128, the SIFT feature vector is created. This vector is normalized and the values which are bigger than 0.2 are thresholded. The thresholded vector is normalized again. Thresholding removes the effect of sudden illumination changes and makes the vector illuminatin invariant as well.

5.2.2 SIFT in Automatic Coin Wear Detection

Apart from being an object detector, SIFT has already been proven to be a powerful texture detector [37][38][39]. Azhar et al. in [39] used SIFT to classify Batik images. They achieved an average of 97.67% accuracy in normal images. However, these studies only cover recognition of well-defined textures. As for the coins, the challenge is that the textural changes are relevant to the other existing quality classes. There is also no rule on where on the coin the change is happening, meaning all the coin's field might have the same amount of wear or the wear can also be concentrated on certain spots.

In coin wear detection, there are several parameters that define the amount of wear on the coins:

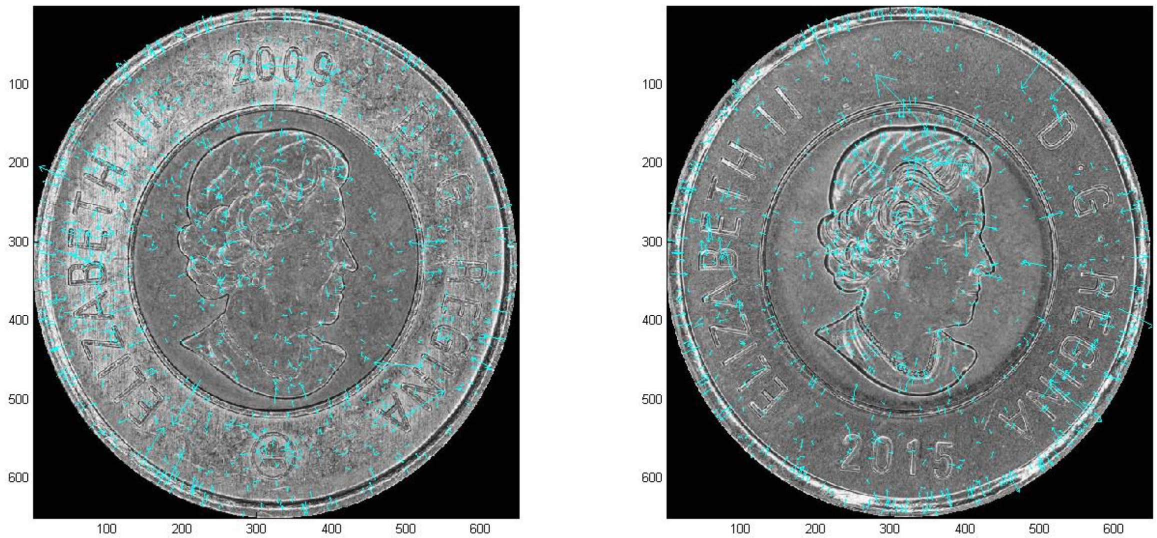
(1) The overall amount of scratches and bumps on the field of the coin

(2) The sharpness of design and legend's edges

Based on the SIFT introduction given in the beginning of this chapter, we claim that these features are capable of recognizing the wear on the coin and in other words, have the unique specifications which make them suitable choice for coin grading.

The SIFT key points include the stable and strong corner points. This is very interesting, since as the coin is worn out, the edges will not be sharp any more. A lot of other edge or gradient based features, will still focus on a worn out edge and therefore, another assistant feature set focused on the field is required to be used along with them. However, the SIFT will be less and less concentrated on the main patterns where the edges are worn out, as the strong corner points are hard to detect. This is also an advantage that the SIFT is not looking for edge key points because of the same reason. In fact, as the coin quality decreases, the amount of overall bumps and scratches on the field increases and therefore, there will be more interesting points to be detected on the field. On the other hand, the concentration of the key points on the design and legend will decrease. When the key points were tested, the practical results were completely align with the theoretical claim given above.

To solve our coin grading problem, we extracted the SIFT key points for 4 randomly selected coins (two uncirculated and two circulated). An average of 2000 to 3000 SIFT key points are detected for each image depending on the level of details. To test the effectiveness of these features for our problem, 50 key points for each coin were selected and visualized on the images. Below, two coin images with different texture qualities are illustrated.



(a) Coin with scratches and bumps

(b) Coin with smooth surface

Figure 5.3: Distribution of SIFT key points on coins with different wear qualities

Glancing at the visualized SIFT key points on the coins illustrated in Figure 5.3, one can easily notice that the key points in Figure 5.3a are more distributed over the field of the coin and shows the defaults in the field texture, whereas the key points in Figure 5.3b are more focused on the corners and edges of the coin's stamp. Moreover, the fact that the design in Figure 5.3a is more worn out gives less key points on the edges and the corners of the design in this coin. Thus we can suggest that if trained well, SIFT key point descriptors can be extremely useful features in classifying the coins with respect to the quality of the wear and the worn out scale.

After making sure that the SIFT descriptors can be beneficial features to discriminate between different coin wear qualities, the algorithm should be tested. Each SIFT key point is described by a histogram of length 128 bins as described in subsection 5.2.1. Hence, the whole image will be shown by a matrix ($n \times 128$) in which n is the number of key points detected in that image. It is obvious that the number of key points in one image is different from the number of key points in another image, since it depends on several parameters such as sharpness of the edges, number of bumps, scratches, and so on. As these parameters vary in different images, we would expect extracting different numbers of key points from each image. On average, a total number of 1500 to 4500 key points were extracted from each image in our dataset depending on the severity of the named parameters. In this step, each image is described by a matrix of key points. To train most of the machine learning algorithms a feature vector is required. Therefore, we shall think of a way to change these feature matrices to meaningful feature vectors for each image.

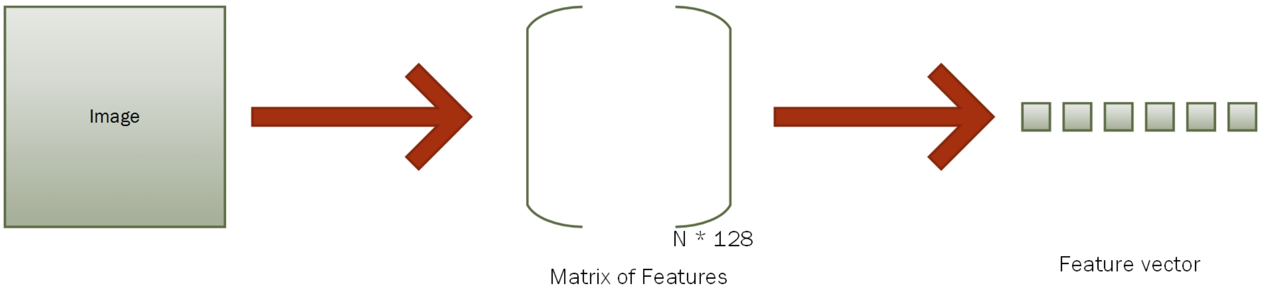


Figure 5.4: Changing the matrix of key points into a feature vector

Two different scenarios were used to make the matrix of features suitable for learning algorithms and the results of each method were investigated.

Method #1

The key points in the training set are labeled based on the image label, meaning that all the key points (i.e. rows in a matrix) extracted from an image inherit the same label as the original image. As discussed in section 5.1.3, the dataset could easily be categorized as uncirculated and circulated coin classes. In this part, the random forest learning algorithm is used due to its reputation for being stable against over-fitting. All the key points extracted from all the images in the training set are then fed to the random forest as the training set. The test set contains the key point descriptors extracted from each test image meaning each row from the extracted key point descriptor matrix behaves as individual test data. As anticipated, the label of each test image is then estimated based on the votes of the labels of its key points.

The result was a training error of 0 a testing error of 38%, which is clearly the case for over-fitting. Therefore, using PCA we reduced the dimension of the data from 128 to 52 in order to deal with the problem. The reduced dimensionality features were still facing over-fitting issue.

Over-fitting and Hoeffding Inequity

Over-fitting is one of the most challenging problems to deal with in machine learning problems. It occurs when the feature size is too large with respect to the data. It is not totally related to the size of the dataset however, size is one important factor. To clarify, if the dataset has a large size but does not require many features to describe the differences among classes, or if the dataset is too small, in both scenarios a large feature set leads to over-fitting. In machine learning problems Inequity 7 describes this complex relation between the size of the feature set and the dataset. In this inequity, M represents the number of features and N is the number of training data examples. In order to prevent the over-fitting, the difference between the training error ($E(in)$) and testing error ($E(out)$) must be smaller than an ϵ value.

Therefore, to prevent over-fitting either:

N is large

, or

M is be in small to compensate for ϵ.

$$P[E(in) - E(out) > \epsilon] \leq 2Me^{-2\epsilon^2 N} \tag{7}$$

Getting back to our dataset, the training set has 100 samples, which is a small number with respect to the study. Therefore, it will generally over-fit for large feature vectors. In our case, through trial and error, we determined that the number of features should be fewer than 8 to be able to control over-fitting. This brought us to the second approach.

Method #2

A common approach in the literature of using the SIFT key point descriptor matrix in a feature vector is to take advantage from *Bag of Visual Words (BoVW)*. BoVW, which is described below, helps to build a meaningful feature vector using the key point descriptors matrix to be fed to any learning algorithm for classification. The common practice in using SIFT descriptors as the features for classification of patterns could be summarized in the following steps:

(1) Extracting the matrix of key point descriptors

(2) Using BoVW to group these key points into clusters (vocabularies)

(3) Create the dictionary of features (bag of features) for each image

(4) Forming a matrix of features using the bag of features created for each image

(5) Using the matrix of features to classify and determine the class of the test images

Bag of Visual Words

In text recognition bag of words is a well-known technique to create a histogram of words for the text, which helps through the process of recognizing the text. Bag of visual words is the application of the same approach in image recognition, which is the reason for calling it bag of visual words instead of bag of words. The technique is presented in the diagram below:

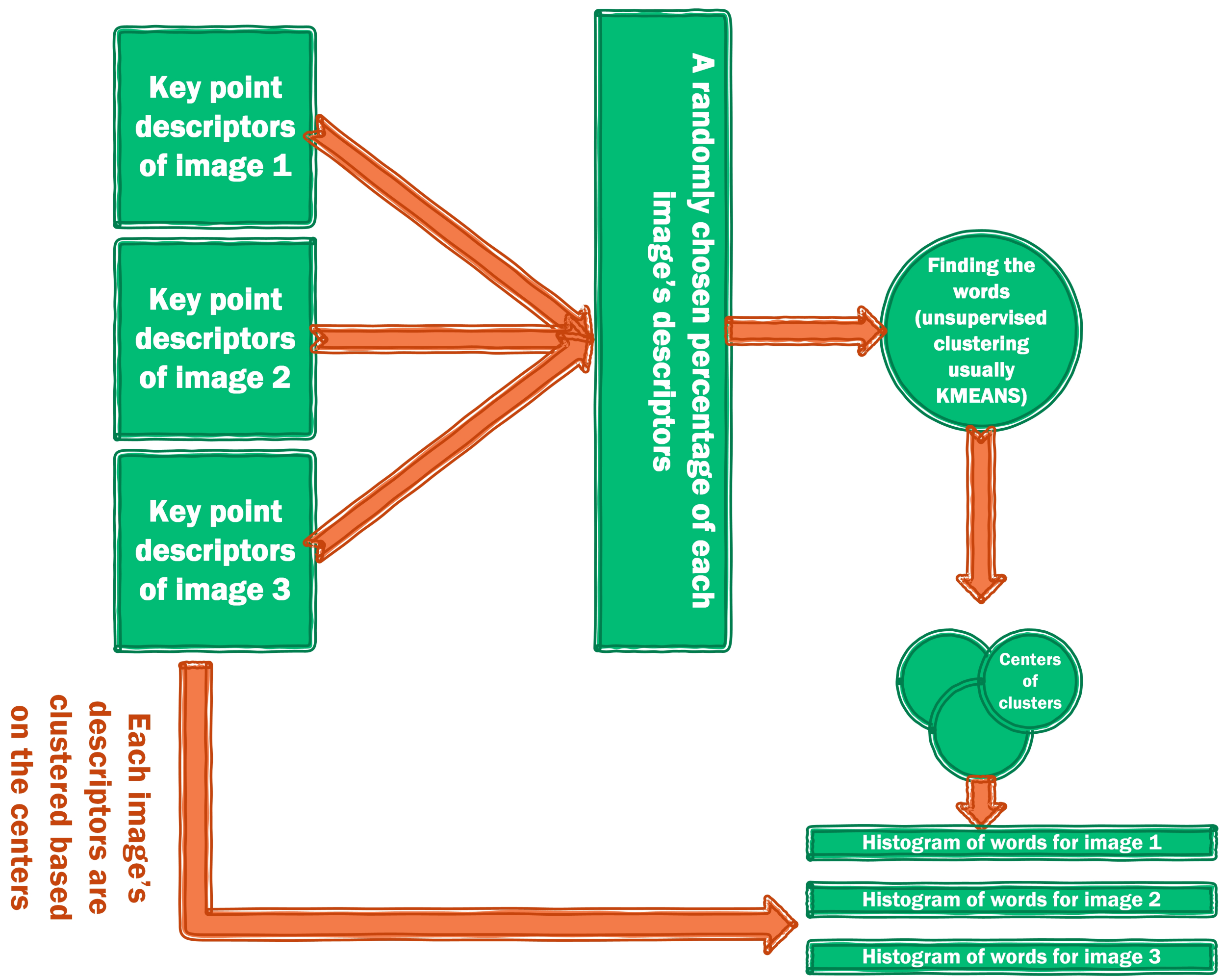


Figure 5.5: Creating the bag of visual words for the each of the three images

As depicted in Figure 5.5 the technique can be summarized in the following steps:

(1) Extracting the SIFT descriptors from all of the images and put them all together in one big SIFT matrix

(2) Grouping the descriptors in a defined number of clusters using an unsupervised clustering technique like k-means

(3) For a given image, after extracting the descriptors, assign each of them to one of the defined clusters in the previous step

(4) Creating a histogram of distribution of the descriptors in the given clusters from second step

It is worth mentioning that in step (2) the clusters are the visual words and we create a dictionary of visual words' counts in this process, which represents the feature vector for the image. BoVW helps reducing the feature vector size to any desired size since the number of clusters is decided more or less by the programmer. In our study, the SIFT descriptors are extracted from all the images in the training set. It has been suggested in the literature [21], not to all the extracted descriptors to construct the visual words. This is because of too much similarity increases the error in finding the correct centers for the clusters. From Equation 7, it is already known in our study that the number of clusters should not exceed 8.

5.3 Classification

Through the previous sections, the images were prepared and the useful information from the images was extracted. The goal of classification is to:

(1) Verify the validity and the reliability of the selected features for the purpose of the study

(2) Complete the computer-aided algorithm designed to perform a task without or with minimal human interference

The choice of a good classifier with respect to the topic and the dataset is another crucial element in each study. One should have a good understanding of the existing classifiers, and their applications in order to choose the ones, which are compatible to the study. Coin grading is about matching the coins with the right existing label (i.e. quality level). This is a supervised study and among the supervised classifiers, Neural Network has been proved to be one of the most powerful ones, when it is trained with a large scale dataset. However, this is not a valid solution for our problem, since the dataset is limited to 129 coins. In the SIFT literature, a combination of SIFT and SVM has given the best results in terms of validity, stability and accuracy [20][21]. There are also so many studies, in which a combination of SIFT and K-NN are used. These two classifiers, are selected to be the main algorithms for this study and will be explained in the following subsections.

For specific occasions, like the one with over-fitting, other classifiers which are known to be stable against the issue, are used. Below, the two selected classifiers for this study are described.

5.3.1 Support Vector Machine (SVM)

Initially, SVM is a binary classifier. The main intention of SVM is to find a hyperplane, with the largest margin between the two classes. Considering x_j as the linearly separable feature points and $y_j \in \{\pm 1\}$ as the labels, the hyperplane is defined as below. β, the weight vector, and b, the bias, are the parameters of the hyperplane.

$$f(X) = Xprime\beta + b = 0 \tag{8}$$

In order to have the maximum margin (i.e. maximum distance to the hyperplane), the following optimization problem is solved [40]:

$$\begin{aligned} &\max_{\beta, b, ||\beta||=1} \quad M \\ &\text{subject to} \quad y_i(x\prime_i \beta + b) \geq M, \; i = 1, \ldots, N. \end{aligned} \tag{9}$$

which is summarized into solving the below Lagrangian optimization problem to find the weight vector β and the bias b:

$$\min_{\beta, b} \quad L(\beta) = 1/2||\beta||^2 - \sum_{i=1}^{N} \alpha_i [y_i(\beta^T x_i + b) - 1] \tag{10}$$

As in any other Lagrangian problem, Equation 10 is solved by taking the derivative of the equation with respect to β and b separately and set it to zero. The obtained amounts for β and b are substituted in 10 and yields a solution to a simpler optimization problem:

$$L_D = \sum_{i=1}^{N} \alpha_i - 1/2 \sum_{i=1}^{N} \sum_{k=1}^{N} \alpha_i \alpha_k y_i y_k x_i^T x_k^T \quad \text{subject to} \quad \alpha_i \geq 0 \tag{11}$$

The above explanation is limited to the linearly separable feature points. For a nonlinearly separable set of points, the data is transformed to a higher dimensional space. In a higher space the dataset is then linearly separable and the formula in 12 could be applied using the transformed data points.

$$L_D = \sum_{i=1}^{N} \alpha_i - 1/2 \sum_{i=1}^{N} \sum_{k=1}^{N} \alpha_i \alpha_k y_i y_k < h(x_i), h(x_k) > \quad \text{subject to} \quad \alpha_i \geq 0 \tag{12}$$

Therefore, the kernel that transforms the data plays the main role in finding a suitable hyperplane. The proper kernel must be selected with respect to the dataset. When the proper kernel is selected, the hyperplane is calculated and transformed to the source space. Therefore, the data is again separated with the largest margin.

For multi-classification problems, SVM can be used by integrating the binary structure into a one against the rest of the classes form. Each time, one of the classes is labeled as 1 and the remaining classes are labeled as -1. The optimum hyperplane for each class against the rest of the classes is designed and the procedure continues for all the existing classes. SVM is recognized for its efficiency and reliability even with a small dataset.

5.3.2 K-Nearest Neighbors (k-NN)

K-nn is a supervised classifier which is classified as a lazy learning algorithm. The whole procedure of learning in this classifier is done in the classification step and for this reason it is a costly algorithm for large scale dataset. K-nn is a powerful algorithm and for small scale dataset where the cost is not an issue, it is a good choice. For each feature point in the testing set, k-nn finds the k nearest neighbors for that point among the training set based on a defined distance measurement. The neighbors vote for the label of that feature point and the majority of the votes defines the label for the test point.

Chapter 6

Results and Discussion

In this chapter, different tests and the results are presented. In the first set of tests, the dataset is classified into two general groups of ”Uncirculated” and ”Circulated” coins to verify the validity of the presented algorithm. In the next set of tests, the labels provided by the coin specialist are used to classify the coins into the three classes of ”Uncirculated”, ”Choice Extremely Fine” and ”Very Fine” quality groups (defined in section 2.2.1). The final results of this section look promising and show the reliability and validity of the method. The performance of the selected machine learning algorithm is also obtained and verified. Through this study, out of 129 coins, 100 are used as the training set and the reminder is used as the testing set. The effects of illumination changes are tested separately.

6.1 Binary Classification Test

In this first set of tests, the coins which were categorized as ”UC” are labeled as ”Uncirculated” and the reminder which includes ”EF+”, ”VF” are labeled as ”Circulated”. Table 6.1 illustrates the distribution of the samples of two existing classes in training and testing sets.

Table 6.1: Distribution of classes in training and testing sets

Type	Number of Uncirculated samples	Number of Circulated samples
Training Set	64	36
Testing Set	20	9

Based on the discussions in chapter 5, section 5.2.2, the SIFT key points are extracted from each of the images in the training set. In order to find the best number of visual words (clusters) in which the key points of training set are best distributed, three tests are done. In each test, a percentage of randomly selected key points (from the whole extracted key points in training set) are used to build 3, 4 and 5 visual words and check how many visual words would give the highest accuracy for the data. The lowest percentage of key points tested for clustering is 50% and the highest is 90%. If choosing less than 50% of the key points, a large portion of information is neglected and going higher than 90% causes too much resemblance in the data which is not a good factor for unsupervised clustering. Clustering the key points into more than 5 groups sometimes failed. The procedure was tested with both SVM and k-nn and the final results are demonstrated in the Tables 6.2 and 6.3 respectively. As can be noticed, for each percentage of used key points in clustering step and number of visual words, the clustering with k-means is tested 3 times. This is due to the fact that for unsupervised clustering, there is always a certain amount of uncertainty in the final result which can be removed by re-doing the test and averaging out the results from different tests. In most cases, repeating the test 3 times, gives the fair result and specially as we are not facing a huge dataset, 3 tests enable us to judge the final result of each step.

Table 6.2: Final accuracy of the algorithm using SVM

% key points used for constructing the VWs	No. Visual Words	Test 1	Test 2	Test 3	Ave. ACC
50%	3	89.6552	93.1034	89.6552	90.8046
	4	89.6552	89.6552	86.2069	88.50577
	5	86.2069	82.7586	79.3103	82.7586
70%	3	89.6552	72.4138	93.1034	85.05747
	4	86.2069	86.2069	86.2069	86.2069
	5	79.3103	75.8621	93.1034	82.7586
90%	3	89.6552	93.1034	93.1034	91.954
	4	89.6552	89.6552	86.2069	88.50577
	5	82.7586	89.6552	89.6552	87.35633

Table 6.3: Final accuracy of the algorithm using 5-NN

% key points used for constructing the VWs	No. Visual Words	Test 1	Test 2	Test 3	Ave. ACC
50%	3	93.1034	93.1034	93.1034	93.1034
	4	93.1034	96.5517	96.5517	95.4023
	5	89.6552	86.2069	89.6552	88.5058
70%	3	89.6552	86.2069	93.1034	89.6552
	4	90.1034	93.1034	96.5517	93.2528
	5	89.6552	86.2069	86.2069	87.3563
90%	3	86.2069	93.1034	93.1034	90.8046
	4	93.1034	96.5517	96.5517	95.4023
	5	86.2069	86.2069	86.2069	86.2069

Based on the discussions in chapter 3, the scale invariant property is already provided for the images. Also, the bigger angular difference among the coin images in the dataset is up to 3 degrees. This is not a noticeable difference to fool SIFT algorithm and therefore, we can confidently say that the tests are scale and rotation invariant. However, the illumination changes in the images are significant. Although SIFT is claimed to be illumination invariant as well as rotation and scale invariant, this property is tested for this study. To do so, the illumination changes in the images are removed using the steps proposed in chapter 3. The SIFT algorithm is then applied to the new images and the results of binary quality classification is illustrated in the Tables 6.4 and 6.5.

Table 6.4: Final accuracy using SVM on images with corrected illumination

% key points used for constructing the VWs	No. Visual Words	Test 1	Test 2	Test 3	Ave. ACC
50%	3	86.2069	86.2069	86.2069	86.2069
	4	96.5517	96.5517	96.5517	96.5517
	5	89.6552	93.1034	96.5517	93.1034
70%	3	89.6552	86.2069	89.6552	88.5058
	4	93.1034	75.8621	96.5517	88.5057
	5	96.5517	93.1034	93.1034	94.2528
90%	3	86.2069	89.6552	89.6552	88.5058
	4	86.2069	96.5517	86.2069	89.6552
	5	93.1034	96.5517	93.1034	94.2528

Table 6.5: Final accuracy using 5-NN on images with corrected illumination

% key points used for constructing the VWs	No. Visual Words	Test 1	Test 2	Test 3	Ave. ACC
50%	3	89.6552	93.1034	93.1034	91.954
	4	96.5517	93.1034	96.5517	95.4023
	5	96.5517	93.1034	93.1034	94.2528
70%	3	93.1034	93.1034	93.1034	93.1034
	4	96.5517	93.1034	96.5517	95.4023
	5	89.6552	89.6552	89.6552	89.6552
90%	3	93.1034	93.1034	93.1034	93.1034
	4	86.2069	96.5517	86.2069	89.6552
	5	89.6552	93.1034	89.6552	90.8046

Comparing the results extracted from the gray-scale images in Tables 6.2 and 6.3 with the ones extracted from the illumination corrected images in Tables 6.4 and 6.5, the accuracy rates calculated for k-nn are more consistent. To find out the best percentage of key points which should participate in the constructing the visual words, the average accuracy of all 9 tests for each percentage of used key points is calculated. The higher accuracy for tests with k-nn in both gray-scale and illumination corrected images goes for using 50% of the key points. Comparing the results of SVM tests, corrected illumination images vote for using 50% of data, whereas the gray-scale images vote for using 90% of key points. This does not only indicate that k-nn results are more consistent but also reveals that using only 50% of all extracted key points from the training set gives us sufficient information to cluster the key points extracted from different parts of the coin and be able to compare them to the relevant key points from other coins. Also, calculating the variance of the accuracies for each test related to a specific number of visual words, shows that on average using 3 clusters is enough for classifying the information. At last, the difference between the results of the illumination corrected images and the results provided in Tables 6.2 and 6.3 is negligible and shows that SIFT features are illumination invariant.

6.1.1 Classification quality measurements

Accuracy is not always enough to show the validity of the classification. It is also dependent on the number of instances in each class. If the size of instances in the existing classes is too much different from each other, the class with a higher number of instances has a greater weight in changing the result of calculated accuracy. Some other factors to evaluate a classifier's performance are investigated below.

Confusion Matrix

Confusion matrix is a tool in machine learning that shows different types of error measurements for a learning system by representing the false positive and the false negative cases. It also shows the performance of the algorithm by representing the true positive and true negative cases all in one matrix. Here, we have selected to show the confusion matrices for the second tests from Tables 6.2 and 6.3 which use 90% of the key points and 4 clusters (defined as test 90.4.2).

SVM	Circulated	Uncirculated
Circulated	7	1
Uncirculated	2	19

(a) SVM test 90.4.2

K-NN	Circulated	Uncirculated
Circulated	8	0
Uncirculated	1	20

(b) K-NN test 90.4.2

Table 6.6: Confusion Matrix of test 90.4.2

Matthew Correlation Coefficient

As illustrated in Table 6.6, the true positive (upper left cell) and true negative (lower right cell) cases in both of the tests are a lot more than false negative (topper right measure) and false positive (lower left) cases. There are several defined quantitative measurements. Each could verify the performance of the learning algorithm from a different aspect and in different situations. The Matthew correlation coefficient is a performance measurement for binary classification when the two classes are very different in size. The Matthews correlation coefficient is defined below:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{13}$$

In the above formulation, TP stands for True Positive, TN for True Negative, FP for False Positive and FN for False Negative. All these variables can be extracted from the confusion matrix. Calculating the MCC for SVM and k-NN using their confusion matrices, MCC for SVM is 0.75 and 0.92 for k-NN. The result shows:

(1) k-NN algorithm has a better performance on our dataset, therefore the accuracy of k-nn cases is more reliable.

(2) Since both SVM and k-NN produce performance results which are a lot higher than 0, in the range between -1 to 1, the algorithm is valid and is proposed to be used for automatic quality detection

6.2 Evaluation Based on Three Levels of Quality

The results in section 6.1 proves the validity of the proposed algorithm to evaluate the coin's quality. The coins in this study are clearly divisible into three quality classes. In order to have a clear decision boundary, the coins are tested to be classified into three different quality classes, which were labeled by the coin specialist (section 2.2.1).

The first class contains the ”UC” labeled samples. The second class includes the ”EF+” labeled samples and the last class contains a combination of ”VF”. The tests results are provided below.

Table 6.7: Final accuracy of the algorithm using SVM with a linear kernel

% key points used for constructing the VWs	No. Visual Words	Test 1	Test 2	Test 3	Ave. ACC
50%	3	89.6552	89.6552	89.6552	89.6552
	4	89.6552	89.6552	89.6552	89.6552
	5	93.1034	93.1034	93.1034	93.1034
70%	3	89.6552	89.6552	89.6552	89.6552
	4	93.1034	89.6552	89.6552	90.8046
	5	93.1034	93.1034	93.1034	93.1034
90%	3	89.6552	89.6552	89.6552	89.6552
	4	93.1034	89.6552	89.6552	90.8046
	5	93.1034	93.1034	93.1034	93.1034

Table 6.8: Final accuracy of the algorithm using SVM with an RBF kernel

% key points used for constructing the VWs	No. Visual Words	Test 1	Test 2	Test 3	Ave. ACC
50%	3	89.6552	89.6552	89.6552	89.6552
	4	96.5517	96.5517	96.5517	96.5517
	5	86.2069	86.2069	86.2069	86.2069
70%	3	86.2069	86.2069	89.6552	87.3581
	4	96.5517	96.5517	96.5517	96.5517
	5	86.2069	86.2069	86.2069	86.2069
90%	3	89.6552	89.6552	89.6552	89.6552
	4	96.5517	96.5517	96.5517	96.5517
	5	93.1034	86.2069	86.2069	88.5075

Table 6.9: Final accuracy of the algorithm using 5-NN

% key points used for constructing the VWs	No. Visual Words	Test 1	Test 2	Test 3	Ave. ACC
50%	3	86.2069	86.2069	86.2069	86.2069
	4	82.7586	89.6552	82.7586	85.0575
	5	89.6552	86.2069	89.6552	88.5067
70%	3	89.6552	89.6552	86.2069	88.5067
	4	89.6552	82.7586	82.7586	85.0575
	5	89.6552	89.6552	82.7586	87.3563
90%	3	86.2069	86.2069	86.2069	86.2069
	4	89.6552	82.7586	89.6552	87.3563
	5	90.1034	89.6552	89.6552	90.8046

In these tests, we used RBf or Gaussian kernel for SVM tests in addition to linear kernel. The reason behind this has been introduced in section 5.1.3. Figure 5.1 shows that the dataset is not easily separable to 3 or 4 classes with linear hypothesis. Calculating the variance of accuracies for each set of 3 tests related to a percentage of key points and a number of visual words, SVM with RBF kernel shows less changes in the final result. Comparing the provided results for k-NN and SVM for classifying in 3 levels of quality, SVM results show better accuracy and more consistency (less variance) through the 3 tests. In other words, when the decision making is more complex, SVM shows a better performance.

Chapter 7

Conclusion

7.1 Contributions and Discussions

This thesis is focused on the feasibility of building an automatic detection of low quality and damaged coins using their digital images. The quality of the coins in numismatic studies are determined based on several factors such as the overall amount of wear, scratches and bumps, color change, struck marks and so on. However, these estimations are not totally solid and can vary based on different perspectives. The aim of an automatic wear estimation is to prevent such judgments while being consistent through the whole procedure.

First, the new dataset and image acquisition software are built from scratch to provided the principal requirements of this thesis. The different aspects of coin photography and challenges have been investigated and the negative impacts of lighting and camera setup on the images have been minimized. The images are cut and scaled automatically and a scale invariant dataset is created using the following techniques: morphological processing, Hough transform, bounding box extraction, and normalization.

Second, as the purpose of the project is to investigate the feasibility and propose an algorithm to do automatic wear detection, the most important factors to estimate the wear quality are determined. The wear quality is mainly estimated based on the overall wear, scratches and bumps. Therefore, the existing texture features have been studied and some have been tested for their performance on this subject. Since illumination changes are not avoidable in coin studies, the study has narrowed down to using SIFT features which are illumination invariant. Other textural features such as local binary patterns which are also used in many studies have failed to reach the goal of this study.

Third, two different approaches to use SIFT features have been tested and the algorithm for wear detection has been proposed and implemented. The algorithm has been tested in several ways to check its validity and feasibility. The challenges with respect to the quantity and variety of the dataset and the over-fitting are recognized and the proper solutions to keep the validity of the overall process are proposed.

At last, to complement the study of this subject, a wear estimation algorithm has been proposed which illustrates that having a good estimation of the coin's quality is mostly related to the overall amount of wear as the main parameter. It also claims that image processing and pattern recognition approaches reveal a great deal of precision which human eye is incapable of distinguishing details in that scale. This responds to the first question of this thesis which is the feasibility of such study and in fact, it indicates the need of having a computerized algorithm for this area of numismatics.

The conclusion of this thesis is summarized below:

- Visual image processing and pattern recognition approach, could give an estimation of the overall wear of the field. However, the quality of the coin also depends on the wear on the design, and legends which more complicated procedures are needed to handle these parameters. Apart from that, the rotation, scale and illumination changes shall be handled in the data preprocessing step. The detected wear pixels should be analysed using a proper scaling system to be able to give a valid estimation on the field wear.

- On the other hand, SIFT features are scale, rotation and illumination invariant which makes them a perfect choice for the numismatic studies. SIFT features are shown to handle the changes in the wear as the coin quality is degraded. They choose the most stable key points which also helps prevent the undesirable effect of noise in the images. All these, helps the different complicated algorithms in the visual image processing approach to be handled automatically using SIFT key points.

- A lot of changes in the patterns of the coins have been handled in this study, using SIFT features, and choosing the right number of clusters for the BoVW algorithm afterwards.

- The SIFT feature extraction process is fast which makes the algorithm to have the capability to be used in real world applications.

7.2 Future Work

As stated in the introduction of this thesis, this study is completely new with respect to the other studies in the field of numismatics. Therefore, some aspects are left for the future work on this subject. The proposition for future work of this subject can be summarized in the following items:

- The factors considered in this thesis are concentrated on the wear and defaults on the coin. However, the other factors such as color changes, eye appeal and so on shall be put into investigation to be added to the proposed algorithm for a complete grading system.

- The dataset used for this study is limited in number and quality variation. A larger dataset with wider variation in the quality shall be used for a better robustness against changes in the coin profile's patterns. If the dataset is large enough, the algorithm is capable to be trained to look for the defaults on the coins and neglect any changes in the main pattern.

- Finally, an algorithm to be completely robust to scale, rotation, illumination and coin's design and legends pattern changes should include a deep learning step. We propose the combination of SIFT key points and deep learning as a new road for discovery in this subject which has the capability to be tested against more complicated applications such as classifying a test coin based on the 70-level quality scale system currently used by the expert numismatists.

References

[1] "Sheldon grading scale 1 - 70," *Austin Rare Coins and Bullio Inc (online brochure). Available at www. austincoins. com/sheldon-grading-scale*, 1989-2016.

[2] J. Martin, "Coin grading and professional third-party grading services," *American Numismatic Association Coin Services (online brochure). Available at www. anacs. com/PDFFiles/ANACS_Brochure. pdf*, 2008.

[3] B.-Y. Feng, K. Sun, P. Atighechian, and C. Y. Suen, "Computer recognition and evaluation of coins," *Handbook of Pattern Recognition and Computer Vision*, pp. 141–158, 2015.

[4] S. Modi and S. Bawa, "Image processing based systems and techniques for the recognition of ancient and modern coins," *International Journal of Computer Applications*, vol. 47, no. 10, pp. 1–5, 2012.

[5] Y. Yadav and A. Sood, "A comparative survey on various coin recognition systems based on image processing," *International Journal of Engineering and Computer Science*, vol. 2, no. 11, pp. 3272–3277, 2013.

[6] M. Reisert, O. Ronneberger, and H. Burkhardt, "An efficient gradient based registration technique for coin recognition," *Proceedings of the Muscle CIS Coin Competition Workshop*, vol. 19, p. 31, 2006.

[7] M. Zaharieva, M. Kampel, and S. Zambanini, *Image Based Recognition of Ancient Coins*, pp. 547–554. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007.

[8] M. Nölle, H. Penz, M. Rubik, K. Mayer, I. Holländer, and R. Granec, “Dagobert A New Coin Recognition and Sorting System,” *Proc. VIIth Digital Image Computing: Techniques and Applications*, pp. 329–338, 2003.

[9] L. J. Van Der Maaten and P. Poon, “Coin-o-matic: A fast system for reliable coin classification,” in *Proc. of the Muscle CIS Coin Competition Workshop, Berlin, Germany*, pp. 7–18, 2006.

[10] R. Huber, H. Ramoser, K. Mayer, H. Penz, and M. Rubik, “Classification of coins using an eigenspace approach,” *Pattern Recognition Letters*, vol. 26, no. 1, pp. 61–75, 2005.

[11] M. Zaharieva, R. Huber, M. Nölle, and M. Kampel, “On ancient coin classification,” *8th International Symposium on Virtual Reality, Archaeology and Cultural Heritage (VAST’07)*, pp. 55–62, 2007.

[12] A. Kavelar, S. Zambanini, and M. Kampel, “Reading the legends of roman republican coins,” *Journal on Computing and Cultural Heritage (JOCCH)*, vol. 7, no. 1, pp. 1–20, 2014.

[13] W. J. Lin and S. S. Jhuo, “Coin recognition based on texture classification on ring and fan areas of the coin image,” *2016 IEEE International Instrumentation and Measurement Technology Conference Proceedings*, pp. 1–6, 2016.

[14] L. Shen, S. Jia, Z. Ji, and W.-S. Chen, “Extracting local texture features for image-based coin recognition,” *IET Image Processing*, vol. 5, no. 5, pp. 394–401, 2011.

[15] D.-M. Tsai and C.-H. Chiang, “Rotation-invariant pattern matching using wavelet decomposition,” *Pattern Recognition Letters*, vol. 23, no. 1, pp. 191–201, 2002.

[16] S. Zambanini, A. Kavelar, and M. Kampel, “Classifying ancient coins by local feature matching and pairwise geometric consistency evaluation,” *Proceedings - International Conference on Pattern Recognition*, vol. 1, no. c, pp. 3032–3037, 2014.

[17] K. Wei, B. He, F. Wang, T. Zhang, and Q. Ding, "A novel method for classification of ancient coins based on image textures," in *Digital Media and its Application in Museum & Heritages, Second Workshop on*, pp. 63–66, IEEE, 2007.

[18] L. Shen, S. Jia, Z. Ji, and W.-S. Chen, "Statictics of gabor features for coin recognition," in *Imaging Systems and Techniques, 2009. IST'09. IEEE International Workshop on*, pp. 295–298, IEEE, 2009.

[19] A. Kavelar, S. Zambanini, and M. Kampel, "Reading ancient coin legends: Object recognition vs. OCR," *CoRR*, vol. abs/1304.7184, no. 1, pp. 1–8, 2013.

[20] H. Anwar, S. Zambanini, and M. Kampel, "A bag of visual words approach for symbols-based coarse-grained ancient coin classification," *CoRR*, vol. abs/1304.6192, no. August 2015, pp. 1–8, 2013.

[21] H. Anwar, S. Zambanini, and M. Kampel, "Coarse-grained ancient coin classification using image-based reverse side motif recognition," *Machine Vision and Applications*, vol. 26, no. 2-3, pp. 295–304, 2015.

[22] J.-P. Wang, Y.-C. Jheng, G.-M. Huang, and J.-H. Chien, "Artificial neural network approach to authentication of coins by vision-based minimization," *Machine vision and applications*, vol. 22, no. 1, pp. 87–98, 2011.

[23] S. Modi and S. Bawa, "Automated coin recognition system using ANN," *CoRR*, vol. abs/1312.6615, no. 4, pp. 975–8887, 2013.

[24] M. Fukumi, S. Omatu, F. Takeda, and T. Kosaka, "Rotation-invariant neural pattern recognition system with application to coin recognition," *IEEE Transactions on neural networks*, vol. 3, no. 2, pp. 272–279, 1992.

[25] A. Khashman, B. Sekeroglu, and K. Dimililer, "Intelligent coin identification system," in

Computer Aided Control System Design, 2006 IEEE International Conference on Control Applications, 2006 IEEE International Symposium on Intelligent Control, 2006 IEEE, pp. 1226–1230, IEEE, 2006.

[26] K. Sun, B. Feng, P. Atighechian, S. Levesque, B. Sinnott, and C. Suen, "Detection of counterfeit coins based on shape and lettering features," in *Proceedings of 28th ISCA International Conference on Computer Applications in Industry and Engineering, San Diego, USA*, pp. 165–170, 2015.

[27] S. Khazaee, M. Sharifi Rad, and C. Y. Suen, "Detection of counterfeit coins based on modeling and restoration of 3d images," in *Computational Modeling of Objects Presented in Images. Fundamentals, Methods and Applications*, pp. 178–193, Springer, 2017.

[28] S. Zambanini and M. Kampel, "Coin data acquisition for image recognition," *Proceedings of the 36th CAA Conference*, pp. 163–170, 2008.

[29] "Coin photography techniques," *TableTop Studio Inc (online brochure). Available at www. tabletopstudio. com/documents/coin photography*, 2008.

[30] M. Nölle, B. Jonsson, and M. Rubik, "Coin Images Seibersdorf - Benchmark," *Smart Systems: High Performance Image Processing*, pp. 1–8, 2004.

[31] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and intelligent laboratory systems*, vol. 2, no. 1-3, pp. 37–52, 1987.

[32] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[33] K. Bagla and B. Bhushan, "A novel approach for face recognition using hybrid sift-svm," in *2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES)*, pp. 1–6, July 2016.

[34] N. Neeru and L. Kaur, “Modified sift descriptors for face recognition under different emotions,” *Journal of Engineering*, vol. 2016, p. 12, 2016.

[35] K. Ogawa, Y. Yamaguchi, Y. Iwamoto, X.-H. Han, and Y.-W. Chen, “Sift-based multi-frame super resolution for 250 million pixel images,” in *Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), International Congress on*, pp. 834–837, IEEE, 2016.

[36] C. Liu, J. Yuen, and A. Torralba, “Sift flow: Dense correspondence across scenes and its applications,” in *Dense Image Correspondences for Computer Vision*, pp. 15–49, Springer, 2016.

[37] D. Erpenbeck, T. Bergen, T. Wittenberg, E. Tannich, C. Wegner, C. Münzenmayer, and M. Benz, “Basic statistics of sift features for texture analysis,” in *Bildverarbeitung für die Medizin 2016*, pp. 98–103, Springer, 2016.

[38] O. Zeglazi, A. Amine, and M. Rziza, “Sift descriptors modeling and application in texture image classification,” in *Computer Graphics, Imaging and Visualization (CGiV), 2016 13th International Conference on*, pp. 265–268, IEEE, 2016.

[39] R. Azhar, D. Tuwohingide, D. Kamudi, and N. Suciati, “Batik image classification using sift feature extraction, bag of features and support vector machine,” *Procedia Computer Science*, vol. 72, pp. 24–30, 2015.

[40] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. New York, NY: Springer New York, 2001.

[41] “The parts of a modern coin,” *GovMINT LLC (online brochure). Available at www. govmint. com/pages/parts-of-modern-coin*, 2017.

Appendix A

Numismatic Terminology and Sheldon Grading Scale

A.1 Numismatic Terminology

Apart from the ”Head” and the ”Tail”, a coin has so many major parts which shall be defined. The terminology used in this thesis is consistent and based on the below definition [41]:

- **Background** - Everything in the image excluding the main object e.g. the coin.
- **Foreground** - The main object of the study e.g. the coin.
- **Obverse** - The ”Head” of the coin.
- **Reverse** - The ”Tail” of the coin.
- **Portrait** - The main part of the design which usually includes the face of a public figure.
- **Field** - The flat area that surrounds coin’s design.
- **Legend** - The principal lettering on the coin usually containing the name of the country, or the name of the figure whose portrait is on the obverse of the coin.

- **Relief** - Generally refers to parts of the coin which are struck in higher positions with respect to the rest of the coin. Coins are built in different degrees of relief.
- **Rim** -The boundary of the coin which is a bit higher than the relief to protect the design from wear.
- **Edge** - The third surface of the coin.
- **Planchet** - The metal disk used to die a coin design on it.

Figure A.1 shows the obverse of a Toonie.

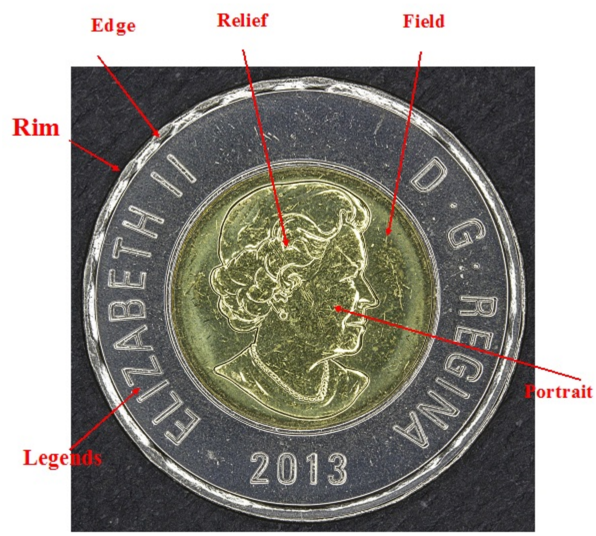


Figure A.1: The important parts of an obverse of a Toonie

A.2 Sheldon Grading Scale

The Sheldon scale is a grading scale devised to facilitate the coin trading and coin collection business. It is initially devised to set price differences for different categories of coins. The categories go further than what is explained in Table A.1. Table A.1 is summarized to cover the wording definition specific to this thesis.

Table A.1: Sheldon Grading Scale [1]

Scale	Grade Label	Specifications
1	Poor (PO)	The coin's type is barely recognizable as it is heavily worn out
2	Fair (FR)	Some small details are distinguishable
3-3.5	Almost Good (AG)	The letters are readable although the coin is heavily worn out
4	Good (G)	The design is visible however in many sections, the coin is worn out flat
6	Choice Good (G+)	Rim and the lettering are complete
8	Very Good (VG)	Slight details are visible although the coin is flat
10	Choice Very Good (VG+)	Slightly clearer design details
12	Fine (F)	Considerable wear with a bold design
20-25	Very Fine (VF)	Moderate wear mostly on relief and high points. The major details are visible
30-35	Choice Very Fine (VF+)	All lettering and main features are sharp with even amount of light wear
40	Extremely Fine (EF)	Light wear on the design with all the features sharp and clear
45	Choice Extremely Fine (EF+)	Light wear on the high points. All the design details are clear. A bit of mint luster is noticeable.
50	Almost Uncirculated (AU)	Slight wear only on high points. Half of mint luster is visible.
60	Uncirculated (UC/MS60)	No trace of wear but may have some contact marks or lack a bit of the mint luster